

Bayesian Modeling of Motion Perception using Dynamical Stochastic Textures

Jonathan Vacher^{1, 4, 5} **Andrew Isaac Meso**^{2, 3}
Laurent U. Perrinet^{2, 5} **and** **Gabriel Peyré**^{1, 6}

¹CEREMADE, Univ. Paris-Dauphine, Paris, FRANCE.

²Institut de Neurosciences de la Timone, Marseille, FRANCE.

³Bournemouth University, Poole, UNITED KINGDOM.

⁴UNIC, Gif-sur-Yvette, FRANCE.

⁵CNRS, FRANCE.

⁶DMA, École Normale Supérieure FRANCE.

Keywords: Dynamic textures, Motion perception, Bayesian Modelling, Stochastic Partial Differential Equations, Psychophysics.

Abstract

A common practice to account for psychophysical biases in vision is to frame them as consequences of a dynamic process relying on optimal inference with respect to a generative model. The present study details the complete formulation of such a generative model intended to probe visual motion perception. It is first derived in a set of axiomatic steps constrained by biological plausibility. We then extend previous contributions by detailing three equivalent formulations of the Gaussian dynamic texture model. First, the composite dynamic textures are constructed by the random aggregation of warped patterns, which can be viewed as 3D Gaussian fields. Second, these textures are cast as solutions to a stochastic partial differential equation (sPDE). This essential step enables real time, on-the-fly, texture synthesis using time-discretized autoregressive processes. It also allows for the derivation of a local motion-energy model, which corresponds to the log-likelihood of the probability density. The log-likelihoods are finally essential for the construction of a Bayesian inference framework. We use the model to probe speed perception in humans psychophysically using zoom-like changes in stimulus spatial frequency content. The likelihood is contained within the generative model and we chose a slow speed prior consistent with previous literature. We then validated the fitting process of the model using synthesized data. The human data replicates previous findings that relative perceived speed is positively biased by spatial frequency increments. The effect cannot be fully accounted for by previous models, but the current prior acting on the spatio-temporal likelihoods has proved necessary in accounting for the perceptual bias.

1 Introduction

1.1 Modeling visual motion perception

A normative explanation for the function of perception is to infer relevant unknown real world parameters from the sensory input, with respect to a generative model [21]. Equipped with some prior knowledge about the nature of neural representation, the modeling representation that emerges corresponds to the *Bayesian brain* hypothesis [28, 12, 7, 27]. This assumes that when given some sensory information S , the brain uses the Bayes theorem :

$$\mathbb{P}_{M|S}(m|s) = \frac{\mathbb{P}_{S|M}(s|m)\mathbb{P}_M(m)}{\mathbb{P}_S(s)}. \quad (1)$$

To estimate the parameters m where the probability distribution function $\mathbb{P}_{S|M}$ is given by the generative model and \mathbb{P}_M represents the prior knowledge. This hypothesis has been well illustrated with the case of motion perception [57]. This uses a Gaussian parameterization of the generative model and a unimodal (Gaussian) prior in order to estimate perceived speed v when observing a visual input I . However, such a Bayesian hypothesis—for instance based on the formalization of unimodal Gaussian prior and likelihood functions—does not always fit with psychophysical results [55, 22]. As such, a major challenge is to refine the definition of generative models so that they are consistent with the widest variety of empirical results.

In fact, the estimation problem inherent to perception is successfully solved, in part, through the definition of an adequate generative model. Probably the simplest generative model to describe visual motion is the luminance conservation equation [2]. It states that luminance $I(x, t)$ for $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ is approximately conserved along trajectories defined as integral lines of a vector field $v(x, t) \in \mathbb{R}^2 \times \mathbb{R}$. The corresponding generative model defines random fields as solutions to the stochastic partial differential equation (sPDE),

$$\langle v, \nabla I \rangle + \frac{\partial I}{\partial t} = W, \quad (2)$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product in \mathbb{R}^2 , ∇I is the spatial gradient of I . To match the spatial scale or frequency statistics of natural scenes (*ie* 1/f amplitude fall-off) or some alternative category of textures, the driving term W is usually defined as a stationary colored Gaussian noise corresponding to the average localized spatio-temporal correlation (which we refer to as spatio-temporal coupling), and is parameterized by a covariance matrix Σ , while the field is usually a constant vector $v(x, t) = v_0$ accounting for a full-field translation with constant speed.

Ultimately, the application of this generative model is essential for probing the visual system, for instance for one seeking to understand how observers might detect motion in a scene. Indeed, as shown by [33, 57], the negative log-likelihood of the probability distribution of the solutions I to the luminance conservation equation (2), on some space-time observation domain $\Omega \times [0, T]$, for some hypothesized constant speed $v(x, t) = v_0$, is proportional to the value of the motion-energy model [2]

$$\int_{\Omega} \int_0^T |\langle v_0, \nabla(K \star I)(x, t) \rangle + \frac{\partial(K \star I)}{\partial t}(x, t)|^2 dt dx \quad (3)$$

where K is the whitening filter corresponding to the inverse square root of Σ , and \star is the convolution operator. Using some prior knowledge about the expected distribution of motions, for instance a preference for slow speeds, a Bayesian formalization can be applied to this inference problem [56, 57].

1.2 Previous Works in Context

Dynamic Texture Synthesis. The model defined in (2) is quite simplistic with respect to the complexity of natural scenes. It is therefore useful here to discuss solutions to generative model problems previously proposed by texture synthesis methods in the computer vision and computer graphics community. Indeed, the literature on the subject of static textures synthesis is abundant (eg [54]). Of particular interest for us is the work of Galerne et al. [17, 16], which proposes a stationary Gaussian model restricted to static textures and provides an equivalent generative model based on Poisson shot noise. Realistic dynamic texture models are however less studied, and the most prominent method is the non-parametric Gaussian auto-regressive (AR) framework of Doretto [11], which has been thoroughly explored [58, 60, 8, 14, 24, 1]. These works generally consists in finding an appropriate low-dimensional feature space in which an

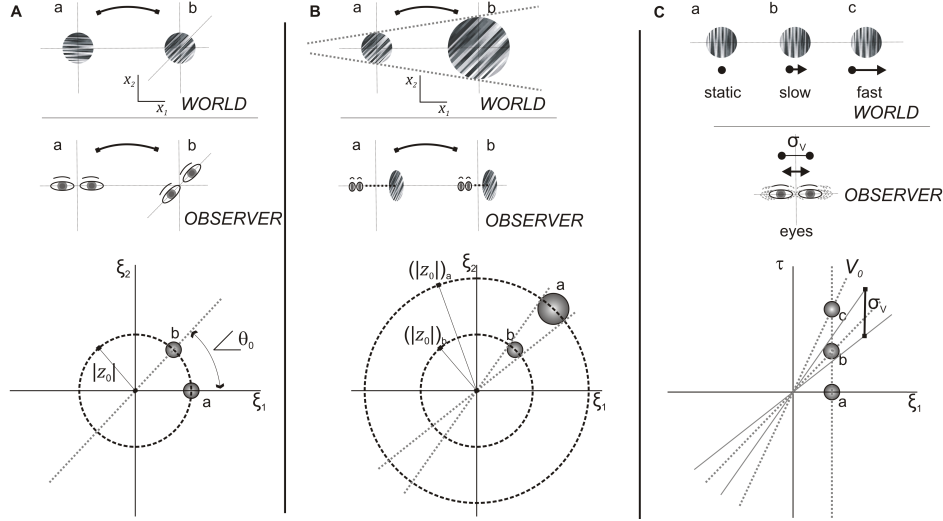


Figure 1: *Parameterization of the class of Motion Clouds stimuli.* The illustration relates the parametric changes in MC with real world (top row) and observer (second row) movements. **(A)** Orientation changes resulting in scene rotation are parameterized through θ as shown in the bottom row where a horizontal a and obliquely oriented b MC are compared. **(B)** Zoom movements, either from scene looming or observer movements in depth, are characterized by scale changes reflected by a scale or frequency term z shown for a larger or closer object b compared to more distant a . **(C)** Translational movements in the scene characterized by V using the same formulation for static (a) slow (b) and fast moving MC, with the variability in these speeds quantified by σ_V . (ξ and τ) in the third row are the spatial and temporal frequency scale parameters. The development of this formulation is detailed in the text.

AR process models the dynamics. Many of these approaches focus on the feature space where decomposition such as Singular Value Decomposition (SVD) and its Higher Order version (HOSVD) [11, 8] has shown their efficiency. In [1], the feature space is the Fourier frequency, and the AR recursion is carried over independently over each frequency, which defines the space-time stationary processes. A similar approach is used in [58] to compute the average of several dynamic texture models. Properties of these AR models are studied by Hyndman [24] where they found that higher order AR processes are able to capture perceptible temporal features. A different approach aims at learning the manifold structure of a given dynamic texture [29] while yet another deals with motion statistics [37]. All these works have in common the will to reproducing the natural spatio-temporal behavior of dynamic textures with rigorous mathematical tools. In addition, our concern is to design a dynamic texture model that is precisely parametrized for experimental purposes in visual neuroscience and psychophysics.

Stochastic Differential Equations (sODE and sPDE). Stochastic Ordinary differential equation (sODE) and their higher dimensional counter-parts, stochastic partial differential equation (sPDE) can be viewed as continuous-time versions of these 1-D or higher dimensional auto-regressive (AR) models. Conversely, AR processes can therefore also be used to compute numerical solutions to these sPDE using finite difference approximations of time derivatives. Informally, these equations can be understood as partial differential equations perturbed by a random noise. The theoretical and numerical study of these sDE is of fundamental interest in fields as diverse as physics and chemistry [52], finance [13] or neuroscience [15]. They allow the dynamic study of complex, irregular and random phenomena such as particle interactions, stocks' or savings' prices, or ensembles of neurons. In psychophysics, sODE have been used to model decision making tasks in which the stochastic variable represents some accumulation of knowledge until the decision is taken, thus providing detailed information about predicted response times [43]. In imaging sciences, sPDE with sparse non-Gaussian driving noise has been proposed as model of natural signals and images [49]. As described above, the simple motion energy model (3) can similarly be demonstrated to rely on the sPDE (2) stochastic model of visual sensory input. This has not previously been presented in a formal way in the literature. One key goal of the current work is to comprehensively describe a parametric family of Gaussian sPDEs which generalize the modeling of moving images (and the corresponding synthesis of visual stimulation) and thus allow for a fine-grained systematic exploration of psychophysical behavior.

Inverse Bayesian inference. Importantly, these dynamic stochastic models are closely related to the likelihood and prior models which serve to infer motion estimates from the dynamic visual stimulation. In order to account for perceptual bias, a now well-accepted methodology in the field of psychophysics is to assume that observers are "ideal observers" and therefore make decisions using optimal statistical inference (typically a maximum-a-posteriori or MAP estimator) which combines this likelihood with some internal prior (1). Several experimental studies use this hypothesis as a justification for the observed perceptual biases by proposing some adjusted likelihood and prior models [12, 7], and more recent works push this idea even further. Observing some

perceptual bias, is it possible to “invert” this forward Bayesian decision making process, and infer the (unknown) internal prior that best fit a set of observed experimental choices made by observers? Following [45], we coined this promising methodology “inverse Bayesian inference”. This is of course an ill-posed, and highly non-linear inverse problem, making it necessary to add constraints on both the prior and the likelihood to make it tractable. For instance [44, 45, 25] impose smoothness constraints in order to be able to locally fit the slope of the prior. Herein, we propose to use visual stimulations generated by the (forward) generative model to challenge this “inverse” Bayesian models. To allow for a simple, yet mathematically rigorous, analysis of this approach within the context of speed discrimination, in the current work we will use a very restricted parametric set of models for the likelihood and priors. This provides a self-consistent approach to test the visual system from the stimulation to the analysis of behavior.

1.3 Contributions

In this paper, we attempt to engender a better understanding of human motion perception by improving generative models for dynamic texture synthesis. From that perspective, we motivate the generation of optimal visual stimulation within a stationary Gaussian dynamic texture model. We develop our current model by extending, mathematically detailing and robustly testing previously introduced dynamic noise textures [39, 40, 51] coined “Motion Clouds”. Our first contribution is a complete axiomatic derivation of the model, seen as a shot noise aggregation of dynamically warped “textons”. Within our generative model, the parameters correspond to average spatial and temporal transformations (*ie* zoom, orientation and translation speed) and associated standard deviations of random fluctuations, as illustrated in Figure 1, with respect to external (objects) and internal (observers) movements. A second contribution is the explicit demonstration of the equivalence between this model and a class of linear sPDEs. This shows that our model is a generalization of the well-known luminance conservation equation 2. This sPDE formulation has two chief advantages: it allows for a real-time synthesis using an AR recurrence and allows one to recast the log-likelihood of the model as a generalization of the classical motion energy model, which in turn is crucial to allow for Bayesian modeling of perceptual biases. Our last contribution follows from the Bayesian approach and is an illustrative application of this model to the psychophysical study of motion perception in humans. This example of the model development constrains the likelihood, which in turn enables a simple fitting procedure to be performed using both an empirical and a larger Monte-Carlo derived synthetic dataset to determine the prior driving the perceptual biases. The code associated to this work is available at <https://jonathanvacher.github.io>.

1.4 Notation

In the following, we will denote $(x, t) \in \mathbb{R}^2 \times \mathbb{R}$ the space/time variable, and $(\xi, \tau) \in \mathbb{R}^2 \times \mathbb{R}$ the corresponding frequency variables. If $f(x, t)$ is a function defined on \mathbb{R}^3 ,

then its Fourier transform is defined as

$$\hat{f}(\xi, \tau) \stackrel{\text{def.}}{=} \int_{\mathbb{R}^2} \int_{\mathbb{R}} f(x, t) e^{-i(\langle x, \xi \rangle + \tau t)} dt dx.$$

For $\xi \in \mathbb{R}^2$, we denote $\xi = \|\xi\|(\cos(\angle \xi), \sin(\angle \xi)) \in \mathbb{R}^2$ its polar coordinates. For a function g defined on \mathbb{R}^2 , we denote $\bar{g}(x) = g(-x)$. In the following, we denote with a capital letter such as A a random variable and a as a realization of A . We note as $\mathbb{P}_A(a)$ the corresponding probability distribution of A .

2 Axiomatic Construction of the Dynamic Textures

Efficient dynamic textures to probe visual perception should be naturalistic yet low-dimensional parametric stochastic models. They should embed meaningful physical parameters (such as the effect of head rotations or whole-field scene movements, see Figure 1) into the local or global dependencies (for instance the covariance) of the random field. In the luminance conservation model (2), the generative model is parameterized by a spatio-temporal coupling encoded in the covariance Σ of the driving noise and the motion flow v_0 .

This localized space-time coupling (e.g. the covariance if one restricts its attention to Gaussian fields) is essential as it quantifies the extent of the spatial integration area as well as the integration dynamics. This is an important issue in neuroscience when considering the implementation of spatio-temporal integration mechanisms from very small to very large scales i.e. going from local to global visual features [38, 3, 9]. In particular, this is crucial to understand the modular sensitivity within the different lower visual areas. In primates for instance, this is evident in the range of spatio-temporal scales of selectivity for generally smaller features observed in the Primary Visual Cortex (V1) and in contrast, ascending the processing hierarchy, for larger features in the Middle Temporal (V5/MT) area. By varying the frequency bandwidth of such dynamic textures, distinct mechanisms for perception and action have been identified in humans [40]. Our goal here is to develop a principled, axiomatic definition of these dynamic textures.

2.1 From Shot Noise to Motion Clouds

We propose a mathematically-sound derivation of a general parametric model of dynamic textures. This model is defined by aggregation, through summation, of a basic spatial “texton” template $g(x)$. The summation reflects a transparency hypothesis, which has been adopted for instance in [17]. While one could argue that this hypothesis is overly simplistic and does not model occlusions or edges, it leads to a tractable framework of stationary Gaussian textures, which has proved useful to model static micro-textures [17] and dynamic natural phenomena [58]. The simplicity of this framework allows for a fine tuning of frequency-based (Fourier) parameterization, which is desirable for the interpretation of psychophysical experiments with respect to underlying spatio-temporal neural sensitivity.

We define a random field as

$$I_\lambda(x, t) \stackrel{\text{def.}}{=} \frac{1}{\sqrt{\lambda}} \sum_{p \in \mathbb{N}} g(\varphi_{A_p}(x - X_p - V_p t)) \quad (4)$$

where $\varphi_a : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a planar warping parameterized by a finite dimensional vector a . The parameters $(X_p, V_p, A_p)_{p \in \mathbb{N}}$ are independent and identically distributed random vectors. They account for the variability in the position of objects or observers and their speed, thus mimicking natural motions in an ambient scene. The set of translations $(X_p)_{p \in \mathbb{N}}$ is a 2-D Poisson point process of intensity $\lambda > 0$. This means that, defining for any measurable A , $C(A) = \# \{p ; X_p \in A\}$, one has that $C(A)$ has a Poisson distribution with mean $\lambda|A|$ (where $|A|$ is the measure of A) and $C(A)$ is independent of $C(B)$ if $A \cap B = \emptyset$.

Intuitively, this model (4) corresponds to a dense mixing of stereotyped, static, textons as in [17]. The originality is two-fold. First, the components of this mixing are derived from the texton by visual transformations φ_{A_p} which may correspond to arbitrary transformations such as zooms or rotations, illustrated in Figure 1. Second, we explicitly model the motion (position X_p and speed V_p) of each individual texton.

In the following, we denote \mathbb{P}_A the common distribution of the i.i.d. $(A_p)_p$, and we denote \mathbb{P}_V the distribution in \mathbb{R}^2 of the speed vectors $(V_p)_p$. Section 2.3 instantiates this model and proposes canonical choices for these variabilities.

The following result shows that the model (4) converges for high point density $\lambda \rightarrow +\infty$ to a stationary Gaussian field and gives the parameterization of the covariance. Its proof follows from a specialization of [16, Theorem 3.1] to our setting.

Proposition 1. *I_λ is stationary with bounded second order moments. Its covariance is $\Sigma(x, t, x', t') = \gamma(x - x', t - t')$ where γ satisfies*

$$\forall (x, t) \in \mathbb{R}^3, \quad \gamma(x, t) = \iiint_{\mathbb{R}^2} c_g(\varphi_a(x - \nu t)) \mathbb{P}_V(\nu) \mathbb{P}_A(a) d\nu da \quad (5)$$

where $c_g = g \star \bar{g}$ is the auto-correlation of g . When $\lambda \rightarrow +\infty$, it converges (in the sense of finite dimensional distributions) toward a stationary Gaussian field I of zero mean and covariance Σ .

This proposition enables us to give a precise definition of a MC.

Definition 1. *A Motion Cloud (MC) is a stationary Gaussian field whose covariance is given by (5).*

Note that, following [18], the convergence result of Proposition 1 could be used in practice to simulate a Motion Cloud I using a high but finite value of λ in order to generate a realization of I_λ . We do not use this approach, and rather rely on the sPDE characterization proved in Section 3, which is well tailored for an accurate and computationally efficient dynamic synthesis.

2.2 Towards “Motion Clouds” for Experimental Purposes

The previous Section provides a theoretical definition of MC 1 that is characterized by $c_g, \varphi_a, \mathbb{P}_A$ and \mathbb{P}_V . The high dimension of these parameters has to be reduced for experimental purposes, therefore it is essential to specify these parameters to have a better control of the covariance γ . We further study this model in the specific case where the warpings φ_a are rotations and scalings (see Figure 1). They account for the characteristic orientations and sizes (or spatial scales) in a scene with respect to the observer. We thus set

$$\forall a = (\theta, z) \in [-\pi, \pi) \times \mathbb{R}_+^*, \quad \varphi_a(x) \stackrel{\text{def.}}{=} zR_{-\theta}(x),$$

where R_θ is the planar rotation of angle θ . We now give some physical and biological motivation underlying our particular choice for the distributions of the parameters. We assume that the distributions \mathbb{P}_Z and \mathbb{P}_Θ of spatial scales z and orientations θ , respectively (see Figure 1), are independent and have densities, thus considering

$$\forall a = (\theta, z) \in [-\pi, \pi) \times \mathbb{R}_+^*, \quad \mathbb{P}_A(a) = \mathbb{P}_Z(z) \mathbb{P}_\Theta(\theta).$$

The speed vector ν is assumed to be randomly fluctuating around a central speed $v_0 \in \mathbb{R}^2$, so that

$$\forall \nu \in \mathbb{R}^2, \quad \mathbb{P}_V(\nu) = \mathbb{P}_{\|\nu - v_0\|}(\|\nu - v_0\|). \quad (6)$$

In order to obtain “optimal” responses to the stimulation (as advocated by [59]) and based on the structure of a standard receptive field of V1, it makes sense to define the texton to be equal to an oriented Gabor which acts as the generic atom

$$g_\sigma(x) = \frac{1}{2\pi} \cos(\langle x, \xi_0 \rangle) e^{-\frac{\sigma^2}{2}\|x\|^2} \quad (7)$$

where σ is the inverse standard deviation and $\xi_0 \in \mathbb{R}^2$ is the spatial frequency. Since the orientation and scale of the texton is handled by the (θ, z) parameters, we can impose without loss of generality the normalization $\xi_0 = (1, 0)$. In the special case where $\sigma \rightarrow 0$, g_σ is a grating of frequency ξ_0 , and the image I is a dense mixture of drifting gratings, whose power-spectrum has a closed form expression detailed in Proposition 2. It is fully parameterized by the distributions $(\mathbb{P}_Z, \mathbb{P}_\Theta, \mathbb{P}_V)$ and the central frequency and speed (ξ_0, v_0) . Note that it is possible to consider any arbitrary textons g , which would give rise to more complicated parameterizations for the power spectrum \hat{g} , but we decided here to stick to the simple asymptotic case of gratings.

Proposition 2. *Consider the texton g_σ , when $\sigma \rightarrow 0$, the Gaussian field $I_\sigma(x, t)$ defined in Proposition 1 converges toward a stationary Gaussian field of covariance having the power-spectrum*

$$\forall (\xi, \tau) \in \mathbb{R}^2 \times \mathbb{R}, \quad \hat{\gamma}(\xi, \tau) = \frac{\mathbb{P}_Z(\|\xi\|)}{\|\xi\|^2} \mathbb{P}_\Theta(\angle \xi) \mathcal{L}(\mathbb{P}_{\|\nu - v_0\|}) \left(-\frac{\tau + \langle v_0, \xi \rangle}{\|\xi\|} \right), \quad (8)$$

where the linear transform \mathcal{L} is such that

$$\forall u \in \mathbb{R}, \quad \mathcal{L}(f)(u) \stackrel{\text{def.}}{=} \int_{-\pi}^{\pi} f(-u / \cos(\varphi)) d\varphi.$$

Proof. We recall the expression (5) of the covariance

$$\forall (x, t) \in \mathbb{R}^3, \quad \gamma(x, t) = \iiint_{\mathbb{R}^2} c_{g_\sigma}(\varphi_a(x - \nu t)) \mathbb{P}_V(\nu) \mathbb{P}_A(a) d\nu da \quad (9)$$

We denote $(\theta, \varphi, z, r) \in \Gamma = [-\pi, \pi]^2 \times \mathbb{R}_+^2$ the set of parameters. Denoting $h(x, t) = c_{g_\sigma}(zR_\theta(x - \nu t))$, one has, in the sense of distributions (taking the Fourier transform with respect to (x, t))

$$\hat{h}(\xi, \tau) = z^{-2} \hat{g}_\sigma(z^{-1}R_\theta(\xi))^2 \delta_{\mathcal{Q}}(\nu) \quad \text{where} \quad \mathcal{Q} = \{\nu \in \mathbb{R}^2; \tau + \langle \xi, \nu \rangle = 0\}.$$

Taking the Fourier transform of (9) and using this computation, the result is that $\hat{\gamma}(\xi, \tau)$ is equal to

$$\int_{\Gamma} \frac{|\hat{g}_\sigma(z^{-1}R_\theta(\xi))|^2}{z^2} \delta_{\mathcal{Q}}(v_0 + r(\cos(\varphi), \sin(\varphi))) \mathbb{P}_\Theta(\theta) \mathbb{P}_Z(z) \mathbb{P}_{\|V-v_0\|}(r) d\theta dz dr d\varphi.$$

Therefore when $\sigma \rightarrow 0$, one has in the sense of distributions

$$|\hat{g}_\sigma(z^{-1}R_\theta(\xi))|^2 \rightarrow \delta_{\mathcal{B}}(\theta, z) \quad \text{where} \quad \mathcal{B} = \{(\theta, z); z^{-1}R_\theta(\xi) = \xi_0\}.$$

Observing that $\delta_{\mathcal{Q}}(\nu) \delta_{\mathcal{B}}(\theta, z) = \delta_{\mathcal{C}}(\theta, z, r)$ where

$$\mathcal{C} = \left\{ (\theta, z, r); z = \|\xi\|, \theta = \angle \xi, r = -\frac{\tau}{\|\xi\| \cos(\angle \xi - \varphi)} - \frac{\|v_0\| \cos(\angle \xi - \angle v_0)}{\cos(\angle \xi - \varphi)} \right\}$$

one obtains the desired formula. \square

Remark 1. Note that the envelope of $\hat{\gamma}$ as defined in (8) is constrained to lie within a cone in the spatio-temporal domain with the apex at zero. This is an important and novel contribution when compared to a classical Gabor. In particular, the bandwidth is then constant around the speed plane or the orientation line with respect to spatial frequency. Basing the generation of the textures on all possible translations, rotations and zooms, we thus provide a principled approach to show that bandwidth should be parametrically scaled with spatial frequency to provide a better model of moving textures.

2.3 Biologically-inspired Parameter Distributions

We now give meaningful specialization for the probability distributions \mathbb{P}_Z , \mathbb{P}_Θ , and $\mathbb{P}_{\|V-v_0\|}$, which are inspired by some known scaling properties of the visual transformations relevant to dynamic scene perception.

Parameterization of \mathbb{P}_Z . First, small, centered, linear movements of the observer along the axis of view (orthogonal to the plane of the scene) generate centered planar zooms of the image. From the linear modeling of the observer's displacement and the subsequent multiplicative nature of zoom, scaling should follow a Weber-Fechner law stating that subjective sensation when quantified is proportional to the logarithm of stimulus intensity. Thus, we choose the scaling z drawn from a log-normal distribution

\mathbb{P}_Z , defined in (10). The bandwidth σ_Z quantifies the variance in the amplitude of zooms of individual textons relative to the characteristic scale z_0 . We thus define

$$\mathbb{P}_Z(z) \propto \frac{\tilde{z}_0}{z} \exp \left(-\frac{\ln \left(\frac{z}{\tilde{z}_0} \right)^2}{2 \ln(1 + \tilde{\sigma}_Z^2)} \right), \quad (10)$$

where \propto means that we ignored the normalizing constant.

In practice, the parameters $(\tilde{z}_0, \tilde{\sigma}_Z)$ are not convenient to manipulate because they have no “physical meaning”. Instead, we use another, more intuitive, parametrization using mode and variance (z_0, σ_Z)

$$z_0 \stackrel{\text{def.}}{=} \operatorname{argmax}_z \mathbb{P}_Z(z) \quad \text{and} \quad \sigma_Z^2 \stackrel{\text{def.}}{=} \mathbb{E}(Z^2) - \mathbb{E}(Z)^2.$$

Once (z_0, σ_Z) are fixed, it is easy to compute the corresponding $(\tilde{z}_0, \tilde{\sigma}_Z)$ to plug into expression (10), simply by solving a polynomial equation (11), as detailed in the following proposition.

Proposition 3. *One has*

$$z_0 = \frac{\tilde{z}_0}{1 + \tilde{\sigma}_Z^2} \quad \text{and} \quad \sigma_Z^2 = \tilde{z}_0^2 \tilde{\sigma}_Z^2 (1 + \tilde{\sigma}_Z^2).$$

Such formula can be inverted by finding the unique positive root of

$$\tilde{\sigma}_Z^2 (1 + \tilde{\sigma}_Z^2)^3 - \frac{\sigma_Z^2}{z_0^2} = 0 \quad \text{and} \quad \tilde{z}_0 = z_0 (1 + \tilde{\sigma}_Z^2). \quad (11)$$

Proof. The primary relations are established using standard calculations from the probability density function \mathbb{P}_Z [26]. The relations (11) follow standard arithmetic. \square

Parametrization of \mathbb{P}_Z by mode and octave bandwidth Differences in perception are often more relevant in a log domain, therefore it is useful to parametrize \mathbb{P}_Z by its mode z_0 and octave bandwidth B_Z which is defined by

$$B_Z \stackrel{\text{def.}}{=} \frac{\ln \left(\frac{z_+}{z_-} \right)}{\ln(2)}$$

where (z_-, z_+) are respectively the successive half-power cutoff frequencies, that is, which verify $\mathbb{P}_Z(z_-) = \mathbb{P}_Z(z_+) = \frac{\mathbb{P}_Z(z_0)}{2}$ with $z_- \leq z_+$.

Proposition 4. *One has*

$$B_Z = \sqrt{\frac{8 \ln(1 + \tilde{\sigma}_Z^2)}{\ln(2)}} \quad \text{and conversely} \quad \tilde{\sigma}_Z = \sqrt{\exp \left(\frac{\ln(2)}{8} B_Z^2 \right) - 1}. \quad (12)$$

Proof. Using the fact that $\mathbb{P}_Z(z_-) = \mathbb{P}_Z(z_+) = \frac{\mathbb{P}_Z(z_0)}{2}$, one shows that $X_+ = \ln\left(\frac{z_+}{z_0}\right)$ and $X_- = \ln\left(\frac{z_-}{z_0}\right)$ are the two roots of the following polynomial (with $X_- \leq X_+$).

$$Q(X) = X^2 + 2\ln(1 + \tilde{\sigma}_Z^2)X - 2\ln(2)\ln(1 + \tilde{\sigma}_Z^2) + \frac{1}{2}\ln(1 + \tilde{\sigma}_Z^2)^2$$

This allows to compute B_Z . □

Through Proposition 4 it is possible to obtain the parametrization of bandwidth prevalent in manipulations used in visual psychophysics experiments.

Parameterization of \mathbb{P}_Θ . Similarly, the texture is perturbed by variations in the global angle θ of the scene: for instance, the head of the observer may roll slightly around its normal position. The von-Mises distribution – as a good approximation of the warped Gaussian distribution around the unit circle – is an adapted choice for the distribution of θ with mean θ_0 and bandwidth σ_Θ ,

$$\mathbb{P}_\Theta(\theta) \propto e^{\frac{\cos(2(\theta-\theta_0))}{4\sigma_\Theta^2}} \quad (13)$$

Parameterization of $\mathbb{P}_{\|V-v_0\|}$. We may similarly consider that the position of the observer is variable in time. On first order approximation, movements perpendicular to the axis of view dominate, generating random perturbations to the global translation v_0 of the image at speed $\nu - v_0 \in \mathbb{R}^2$. These perturbations are for instance described by a Gaussian random walk: take for instance tremors, which are constantly jittering, small (≤ 1 deg) movements of the eye. This justifies the choice of a radial distribution (6) for \mathbb{P}_V . This radial distribution $\mathbb{P}_{\|V-v_0\|}$ is thus selected as a bell-shaped function of width σ_V , and we choose here a Gaussian function for simplicity

$$\mathbb{P}_{\|V-v_0\|}(r) \propto e^{-\frac{r^2}{2\sigma_V^2}}. \quad (14)$$

Note that, as detailed in Section 3.2 a slightly different bell-function (with a more complicated expression) should be used to obtain an exact equivalence with the sPDE discretization.

Putting everything together. Plugging these expressions (10), (13) and (14) into the definition (8) of the power spectrum of the motion cloud, one obtains a parameterization which shares similarities with the one originally introduced in [40].

The following table recaps the parameters of the biologically-inspired MC models. It is composed of the central parameters (v_0) for the speed, (θ_0) for orientation and (z_0) for the frequency modulus, as well as corresponding “dispersion” parameters ($\sigma_V, \sigma_\Theta, B_Z$) which account for the typical deviation around these centers.

	Speed	Freq. orient.	Freq. amplitude
(mean, dispersion)	(v_0, σ_V)	(θ_0, σ_Θ)	(z_0, B_Z)

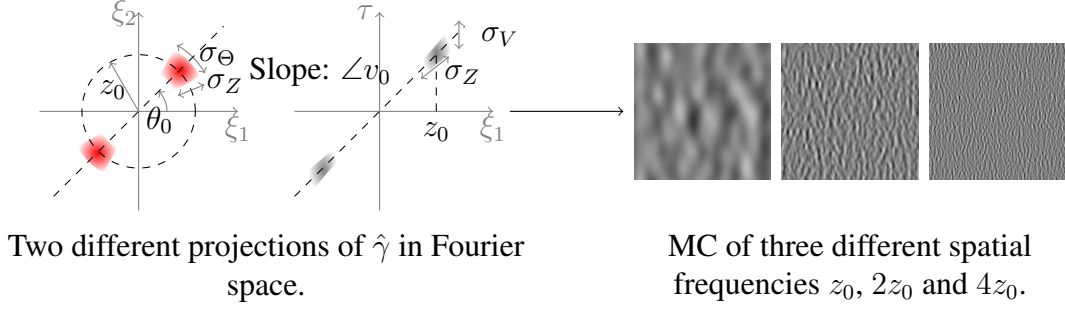


Figure 2: Graphical representation of the covariance $\hat{\gamma}$ (left) —note the cone-like shape of the envelopes— and an example of synthesized frames for three different spatial frequency (right).

Figure 2 shows graphically the influence of these parameters on the shape of the MC power spectrum $\hat{\gamma}$.

We show in Figure 3 two examples of such stimuli for different spatial frequency bandwidths. In particular, by tuning this bandwidth, in previous studies it has been possible to dissociate its respective role in action and perception [40]. Using this formulation to extend the study of visual perception to other dimensions, such as orientation or speed bandwidths, should provide a means to systematically titrate their respective role in motion integration and obtain essential novel data.

3 sPDE Formulation and Synthesis Algorithm

In this section, we show that the MC model (Definition 1) can equally be described as the stationary solution of a stochastic partial differential equation (sPDE). This sPDE formulation is important since we aim to deal with dynamic stimulation, which should be described by a causal equation which is local in time. This is crucial for numerical simulations, since this allows us to perform real-time synthesis of stimuli using an auto-regressive time discretization. This is a significant departure from previous Fourier-based implementation of dynamic stimulation [39, 40]. Moreover, this is also important to simplify the application of MC inside a Bayesian model of psychophysical experiments (see Section 4). In particular, the derivation of an equivalent sPDE model exploits a spectral formulation of MCs as Gaussian Random fields. The full proof along with the synthesis algorithm follows.

To be mathematically correct, all the sPDE in this article are written in the sense of generalized stochastic processes (GSP) which are to stochastic processes what generalized functions are to functions. This allows the consideration of linear transformations of stochastic processes like differentiation or Fourier transforms as for generalized functions. We refer to [50] for a recent use of GSP and to [19] for the foundation of the theory. The connection between GSP and stochastic processes has been described by previous work [30]

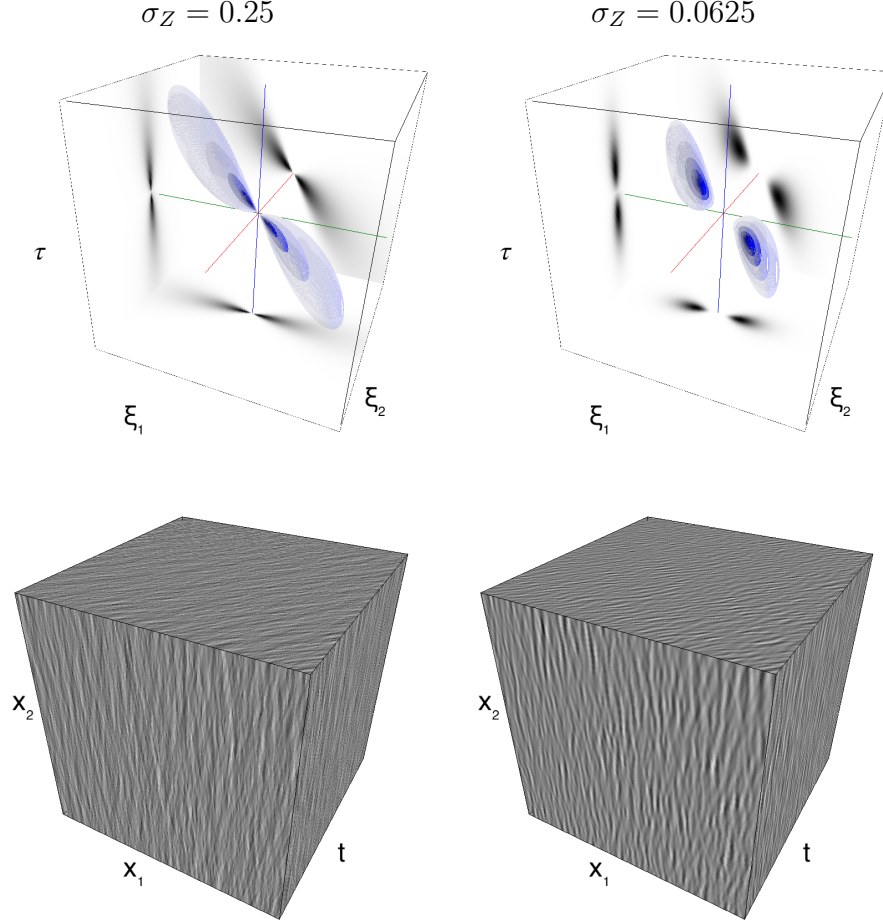


Figure 3: Comparison of the broadband (left) vs. a narrowband (right) stimulus. Two instances (left and right columns) of two motions clouds having the same parameters except the frequency bandwidths σ_Z , which were different. The top column displays iso-surfaces of $\hat{\gamma}$ in the form of enclosing volumes at different energy values with respect to the peak amplitude of the Fourier spectrum. The bottom column shows an isometric view of the faces of a movie cube, which is a realization of the random field I . The first frame of the movie lies on the $(x_1, x_2, t = 0)$ spatial plane. The Motion Cloud with the broadest bandwidth is often thought to best represent stereotyped natural stimuli, since, it similarly contains a broad range of frequency components.

3.1 Dynamic Textures as Solutions of sPDE

In the following, we first restrict our attention to the case $v_0 = 0$ in order to define a simple sPDE, and then detail the general case.

sPDE without global translation, $v_0 = 0$. We first give the definition of a sPDE cloud I making use of another cloud I_0 without translation speed.

Definition 2. For a given spatial covariance Σ_W , 2-D spatial filters (α, β) and a trans-

lation speed $v_0 \in \mathbb{R}^2$, a sPDE cloud is defined as

$$I(x, t) \stackrel{\text{def.}}{=} I_0(x - v_0 t, t). \quad (15)$$

where I_0 is a stationary Gaussian field satisfying for all (x, t)

$$\mathcal{D}(I_0) = \frac{\partial W}{\partial t} \quad \text{where} \quad \mathcal{D}(I_0) \stackrel{\text{def.}}{=} \frac{\partial^2 I_0}{\partial t^2} + \alpha \star \frac{\partial I_0}{\partial t} + \beta \star I_0 \quad (16)$$

where the driving noise $\frac{\partial W}{\partial t}$ is white in time (i.e. corresponding to the temporal derivative of a Brownian motion in time) and has a 2-D stationary covariance σ_W in space and \star is the spatial convolution operator.

The random field I_0 solving (16) thus corresponds to a sPDE cloud with no translation speed, $v_0 = 0$. The filters (α, β) parameterizing this sPDE cloud aim at enforcing an additional correlation in time of the model. Section 3.2 explains how to choose $(\alpha, \beta, \sigma_W)$ so that these sPDE clouds, which are stationary solutions of (16), have the power spectrum given in (8) (in the case that $v_0 = 0$), i.e. are motion clouds.

Defining a causal equation that is local in time is crucial for numerical simulation (as explained in Section 3.3) but also to simplify the application of MC inside a Bayesian model of psychophysical experiments (see Section 4.3).

The sPDE equation (16) corresponds to a set of independent stochastic ODEs over the spatial Fourier domain, which reads, for each frequency ξ ,

$$\forall t \in \mathbb{R}, \quad \frac{\partial^2 \hat{I}_0(\xi, t)}{\partial t^2} + \hat{\alpha}(\xi) \frac{\partial \hat{I}_0(\xi, t)}{\partial t} + \hat{\beta}(\xi) \hat{I}_0(\xi, t) = \hat{\sigma}_W(\xi) \hat{w}(\xi, t) \quad (17)$$

where $\hat{I}_0(\xi, t)$ denotes the Fourier transform with respect to the spatial variable x only. Here, $\hat{\sigma}_W(\xi)^2$ is the spatial power spectrum of $\frac{\partial W}{\partial t}$, which means that

$$\Sigma_W(x, y) = c(x - y) \quad \text{where} \quad \hat{c}(\xi) = \hat{\sigma}_W^2(\xi). \quad (18)$$

Here $\hat{w}(\xi, t) \sim \mathcal{CN}(0, 1)$ and w is a white noise in space and time.

While the equation (17) should hold for all time $t \in \mathbb{R}$, the construction of stationary solutions (hence sPDE clouds) of this equation is obtained by solving the sODE (17) forward for time $t > t_0$ with arbitrary boundary conditions at time $t = t_0$, and letting $t_0 \rightarrow -\infty$. This is consistent with the numerical scheme detailed in Section 3.3.

While it is beyond the scope of this paper to study theoretically the equation (16), one can show the existence and uniqueness results of stationary solutions for this class of sPDE under stability conditions on the filters (α, β) (see for instance [49, 5]) that are automatically satisfied for the particular case of Section 3.2.

Theorem 1. *If $(\hat{\alpha}, \hat{\beta})$ are non-negative and $\frac{\hat{\sigma}_W^2}{\hat{\alpha}\hat{\beta}} \in L^1$, then Equation (16) has a unique causal and stationary solution, i.e. it defines uniquely the distribution of a sPDE cloud.*

Proof. Consider (17), the Fourier transform of (16) which has causal and stationary solutions (see the general case of Levy-driven sPDE, Theorem 3.3 in [5]). Hence $\frac{\hat{\sigma}_W}{\hat{\alpha}\hat{\beta}} \in L^1$, these solutions have an integrable spatial power spectrum. Then, one could take their inverse Fourier transform and get the solution which is unique by construction. \square

Remark 2. There are different ways to define uniqueness of solution for sPDE. In Theorem 1, uniqueness has to be understood in terms of sample path, meaning that two solutions (X, \tilde{X}) of Equation (16) verifies $\mathbb{P}(\forall t \in \mathbb{R}, X_t = \tilde{X}_t) = 1$. This notion of uniqueness is strong and it implies uniqueness in distribution meaning that X and \tilde{X} have the same law.

sPDE with global translation. The easiest way to define and synthesize a sPDE cloud I with non-zero translation speed v_0 is to first define I_0 solving (17) and then translating it with constant speed using (15). An alternative way is to derive the sPDE satisfied by I , as detailed in the following proposition. This is useful to define motion energy in Section 4.3.

Proposition 5. *The MCs noted I with speed parameters $(\alpha, \beta, \Sigma_W)$ and translation speed v_0 are the stationary solutions of the sPDE*

$$\mathcal{D}(I) + \langle \mathcal{G}(I), v_0 \rangle + \langle \mathcal{H}(I)v_0, v_0 \rangle = \frac{\partial W}{\partial t} \quad (19)$$

where \mathcal{D} is defined in (16), $\partial_x^2 I$ is the Hessian of I (second order spatial derivative), where

$$\mathcal{G}(I) \stackrel{\text{def.}}{=} \alpha \star \nabla_x I + 2\partial_t \nabla_x I \quad \text{and} \quad \mathcal{H}(I) \stackrel{\text{def.}}{=} \nabla_x^2 I. \quad (20)$$

Proof. This follows by computing the derivative in time of the warping equation (15), denoting $y \stackrel{\text{def.}}{=} x + v_0 t$

$$\begin{aligned} \partial_t I_0(x, t) &= \partial_t I(y, t) + \langle \nabla I(y, t), v_0 \rangle, \\ \partial_t^2 I_0(x, t) &= \partial_t^2 I(y, t) + 2\langle \partial_t \nabla I(y, t), v_0 \rangle + \langle \partial_x^2 I(y, t)v_0, v_0 \rangle \end{aligned}$$

and plugging this into (16) after remarking that the distribution of $\frac{\partial W}{\partial t}(x, t)$ is the same as the distribution of $\frac{\partial W}{\partial t}(x - v_0 t, t)$. \square

3.2 Equivalence between the spectral and sPDE formulations

Since both MCs and sPDE clouds are obtained by a uniform translation with speed v_0 of a motionless cloud, we can restrict without loss of generality our analysis to the case $v_0 = 0$.

In order to relate MCs to sPDE clouds, equation (17) makes explicit that the functions $(\hat{\alpha}(\xi), \hat{\beta}(\xi))$ should be chosen in order for the temporal covariance of the resulting process to be equal (or at least to approximate well) the temporal covariance appearing in (8). This covariance should be localized around 0 and be non-oscillating. It thus makes sense to constrain $(\hat{\alpha}(\xi), \hat{\beta}(\xi))$ for the corresponding ODE (17) to be critically damped, which corresponds to imposing the following relationship

$$\forall \xi, \quad \hat{\alpha}(\xi) = \frac{2}{\hat{\nu}(\xi)} \quad \text{and} \quad \hat{\beta}(\xi) = \frac{1}{\hat{\nu}^2(\xi)}$$

for some relaxation step size $\hat{\nu}(\xi)$. The model is thus solely parameterized by the noise variance $\hat{\sigma}_W(\xi)$ and the characteristic time $\hat{\nu}(\xi)$.

The following proposition shows that the sPDE cloud model (16) and the motion cloud model (8) are identical for an appropriate choice of function $\mathbb{P}_{\|V-v_0\|}$.

Proposition 6. *When considering*

$$\forall r > 0, \quad \mathbb{P}_{\|V-v_0\|}(r) = \mathcal{L}^{-1}(h)(r/\sigma_V) \quad \text{where} \quad h(u) = (1 + u^2)^{-2} \quad (21)$$

where \mathcal{L} is defined in (8), equation (16) admits a solution I which is a stationary Gaussian field with power spectrum (8) when setting

$$\hat{\sigma}_W^2(\xi) = \frac{1}{\hat{\nu}(\xi)^3 \|\xi\|^2} \mathbb{P}_Z(\|\xi\|) \mathbb{P}_\Theta(\angle \xi), \quad \text{and} \quad \hat{\nu}(\xi) = \frac{1}{\sigma_V \|\xi\|}. \quad (22)$$

Proof. For this proof, we denote I^{MC} the motion cloud defined by (8), and I a stationary solution of the sPDE defined by (16) which exists according to Theorem 1 because $\hat{\sigma}_W^2 \hat{\nu}^3 \in L^1$, indeed \mathbb{P}_Z and \mathbb{P}_Θ are probability distributions and $\xi \mapsto \frac{1}{\|\xi\|^2}$ does not change the continuity at 0. We aim to show that under the specification (22), they have the same covariance. This is equivalent to showing that $I_0^{\text{MC}}(x, t) = I^{\text{MC}}(x + ct, t)$ has the same covariance as I_0 . For any fixed ξ , equation (17) admits a unique stationary solution $\hat{I}_0(\xi, \cdot)$ (Theorem 1) which is a stationary Gaussian process of zero mean and with a covariance which is $\hat{\sigma}_W^2(\xi) r \star \bar{r}$ where r is the impulse response (i.e. taking formally $a = \delta$) of the ODE $r'' + 2r'/u + r''/u^2 = a$ where we denoted $u = \hat{\nu}(\xi)$. This impulse response can be shown to be $r(t) = te^{-t/u} \mathbb{1}_{\mathbb{R}^+}(t)$. The covariance of $\hat{I}_0(\xi, \cdot)$ is thus, after some computation, equal to $\hat{\sigma}_W^2(\xi) r \star \bar{r} = \hat{\sigma}_W^2(\xi) h(\cdot/u)$ where $h(t) \propto (1 + |t|)e^{-|t|}$. Taking the Fourier transform of this equality, the power spectrum $\hat{\gamma}_0$ of I_0 thus reads

$$\hat{\gamma}_0(\xi, \tau) = \hat{\sigma}_W^2(\xi) \hat{\nu}(\xi)^3 \hat{h}(\hat{\nu}(\xi)\tau) \quad \text{where} \quad \hat{h}(u) = \frac{1}{(1 + u^2)^2} \quad (23)$$

and where it should be noted that this function h is the same as the one introduced in (21). The covariance γ^{MC} of I^{MC} and γ_0^{MC} of I_0^{MC} are related by the relation

$$\hat{\gamma}_0^{\text{MC}}(\xi, \tau) = \hat{\gamma}^{\text{MC}}(\xi, \tau - \langle \xi, v_0 \rangle) = \frac{1}{\|\xi\|^2} \mathbb{P}_Z(\|\xi\|) \mathbb{P}_\Theta(\angle \xi) \hat{h}\left(-\frac{\tau}{\sigma_V \|\xi\|}\right). \quad (24)$$

where we used the expression (8) for $\hat{\gamma}^{\text{MC}}$ and the value of $\mathcal{L}(\mathbb{P}_{\|V-v_0\|})$ given by (21). Condition (22) guarantees that expression (23) and (24) coincide, and thus $\hat{\gamma}_0 = \hat{\gamma}_0^{\text{MC}}$. \square

Expression for $\mathbb{P}_{\|V-v_0\|}$. Equation (21) states that in order to obtain a perfect equivalence between the MC defined by (8) and by (16), the function $\mathcal{L}^{-1}(h)$ has to be well-defined. It means we need to compute the inverse of the transform of the linear operator \mathcal{L}

$$\forall u \in \mathbb{R}, \quad \mathcal{L}(f)(u) = 2 \int_0^{\pi/2} f(-u/\cos(\varphi)) d\varphi.$$

to the function h . The following proposition gives a closed-form expression for this function, and shows in particular that it is a function in $L^1(\mathbb{R})$, i.e. it has a finite integral, which can be normalized to unity to define a density distribution. Figure 4 shows a graphical display of that distribution.

Proposition 7. *One has*

$$\mathcal{L}^{-1}(h)(u) = \frac{2 - u^2}{\pi(1 + u^2)^2} - \frac{u^2(u^2 + 4)(\log(u) - \log(\sqrt{u^2 + 1} + 1))}{\pi(u^2 + 1)^{5/2}}.$$

In particular, one has

$$\mathcal{L}^{-1}(h)(0) = \frac{2}{\pi} \quad \text{and} \quad \mathcal{L}^{-1}(h)(u) \sim \frac{1}{2\pi u^3} \quad \text{when} \quad u \rightarrow +\infty.$$

Proof. The variable substitution $x = \cos(\varphi)$ can be used to rewrite (3.2) as

$$\forall u \in \mathbb{R}, \quad \mathcal{L}(h)(u) = 2 \int_0^1 h\left(-\frac{u}{x}\right) \frac{x}{\sqrt{1 - x^2}} \frac{dx}{x}.$$

In such a form, we recognize a Mellin convolution which could be inverted by the use of Mellin convolution table [34]. \square

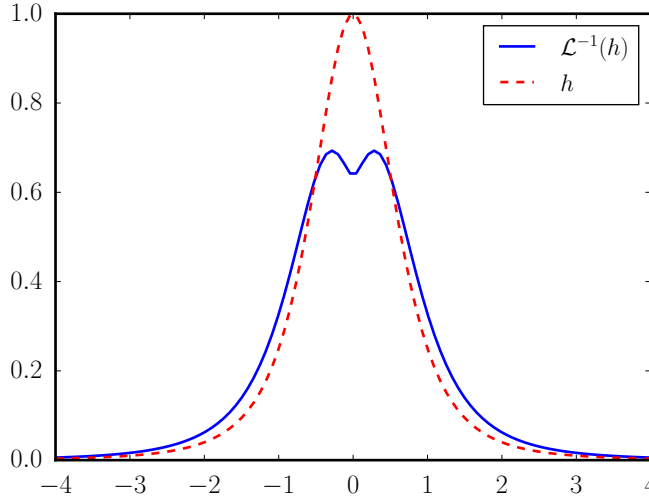


Figure 4: Functions h and $\mathcal{L}^{-1}(h)$.

3.3 AR(2) Discretization of the sPDE

Most previous works for Gaussian texture synthesis (such as [17] for static and [39, 40] for dynamic textures) have used a global Fourier-based approach and the explicit power spectrum expression (8). The main drawbacks of such an approach are: (i) it introduces an artificial periodicity in time and thus can only be used to synthesize a finite number of frames; (ii) these frames must be synthesized at once, before the stimulation, which prevents real-time synthesis; (iii) the discrete computational grid may introduce artifacts, in particular when one of the included frequencies is of the order of the discretization step or when a bandwidth is too small.

To address these issues, we follow the previous works of [11, 58] and make use of an auto-regressive (AR) discretization of the sPDE (16). In contrast with these previous

works, we use a second order AR(2) regression (in place of a first order AR(1) model). Using higher order recursions is crucial to make the output consistent with the continuous formulation (16). Indeed, numerical simulations show that AR(1) iterations lead to unacceptable temporal artifacts: in particular, the time correlation of AR(1) random fields typically decays too fast in time.

AR(2) synthesis without global translation, $v_0 = 0$. The discretization computes a (possibly infinite) discrete set of 2-D frames $(I_0^{(\ell)})_{\ell \geq \ell_0}$ separated by a time step Δ , and we approach at time $t = \ell\Delta$ the derivatives as

$$\frac{\partial I_0(\cdot, t)}{\partial t} \approx \Delta^{-1}(I_0^{(\ell)} - I_0^{(\ell-1)}) \quad \text{and} \quad \frac{\partial^2 I_0(\cdot, t)}{\partial t^2} \approx \Delta^{-2}(I_0^{(\ell+1)} + I_0^{(\ell-1)} - 2I_0^{(\ell)}),$$

which leads to the following explicit recursion

$$\forall \ell \geq \ell_0, \quad I_0^{(\ell+1)} = (2\delta - \Delta\alpha - \Delta^2\beta) \star I_0^{(\ell)} + (-\delta + \Delta\alpha) \star I_0^{(\ell-1)} + \Delta^2 W^{(\ell)}, \quad (25)$$

where δ is the 2-D Dirac distribution and where $(W^{(\ell)})_\ell$ are i.i.d. 2-D Gaussian field with distribution $\mathcal{N}(0, \Sigma_W)$, and $(I_0^{(\ell_0-1)}, I_0^{(\ell_0-1)})$ can be arbitrary initialized.

One can show that when $\ell_0 \rightarrow -\infty$ (to allow for a long enough “warmup” phase to reach approximate time-stationarity) and $\Delta \rightarrow 0$, then I_0^Δ defined by interpolating $I_0^\Delta(\cdot, \Delta\ell) = I_0^{(\ell)}$ converges (in the sense of finite dimensional distributions) toward a solution I_0 of the sPDE (16). Here we choose to use the standard finite difference however we refer to [48, 4] for more advanced discretization. We implemented the recursion (25) by computing the 2-D convolutions with FFT’s on a GPU, which allows us to generate high resolution videos in real time, without the need to explicitly store the synthesized video.

AR(2) synthesis with global translation. The easiest way to approximate a sPDE cloud using an AR(2) recursion is to simply apply formula (15) to $(I_0^{(\ell)})_\ell$ as defined in (25), that is, to define

$$I^{(\ell)}(x) \stackrel{\text{def.}}{=} I_0^{(\ell)}(x - v_0\Delta\ell).$$

A second alternative approach would be to directly discretized the sPDE (19). We did not use this approach because it requires the discretization of spatial differential operators \mathcal{G} and \mathcal{H} , and is hence less stable. A third, somehow hybrid, approach, is to apply the spatial translations to the AR(2) recursion, and define the following recursion

$$I^{(\ell+1)} = \mathcal{U}_{v_0} \star I^{(\ell)} + \mathcal{V}_{v_0} \star I^{(\ell-1)} + \Delta^2 W^{(\ell)}, \quad (26)$$

$$\text{where} \quad \begin{cases} \mathcal{U}_{v_0} \stackrel{\text{def.}}{=} (2\delta - \Delta\alpha - \Delta^2\beta) \star \delta_{-\Delta v_0}, \\ \mathcal{V}_{v_0} \stackrel{\text{def.}}{=} (-\delta + \Delta\alpha) \star \delta_{-2\Delta v_0}, \end{cases} \quad (27)$$

where δ_s indicates the Dirac at location s , so that $(\delta_s \star I)(x) = I(x - s)$ implements the translation by s . Numerically, it is possible to implement (26) over the Fourier domain,

$$\hat{I}^{(\ell+1)}(\xi) = \hat{\mathcal{U}}_{v_0}(\xi) \hat{I}^{(\ell)}(\xi) + \hat{\mathcal{V}}_{v_0}(\xi) \hat{I}^{(\ell-1)}(\xi) + \Delta^2 \hat{\sigma}_W(\xi) \hat{w}^{(\ell)}(\xi),$$

$$\text{where} \quad \begin{cases} \hat{\mathcal{U}}_{v_0}(\xi) \stackrel{\text{def.}}{=} (2 - \Delta\hat{\alpha}(\xi) - \Delta^2\hat{\beta}(\xi))e^{-i\Delta v_0\xi}, \\ \hat{\mathcal{Q}}_{v_0}(\xi) \stackrel{\text{def.}}{=} (-1 + \Delta\hat{\alpha}(\xi))e^{-2i\Delta v_0\xi}, \end{cases}$$

and where $w^{(\ell)}$ is a 2-D white noise.

4 An Empirical Study of Visual Speed Discrimination

To exploit the useful parametric transformation features of our MC model and provide a generalizable proof of concept based on motion perception, we consider here the problem of judging the relative speed of moving dynamical textures. The overall aim is to characterize the impact of both average spatial frequency and average duration of temporal correlations on perceptual speed estimation based on the empirical evidence.

4.1 Methods

The task was to discriminate the speed $v \in \mathbb{R}$ of a MC stimuli moving with a horizontal central speed $\mathbf{v} = (v, 0)$. We assign as independent experimental variable the most represented spatial frequency z_0 , that we denote in the following z for easier reading. The other parameters are set to the following values

$$\sigma_V = \frac{1}{t^* z}, \quad \theta_0 = \frac{\pi}{2}, \quad \sigma_\Theta = \frac{\pi}{12}.$$

Note that σ_V is thus dependent of the value of z to ensure that $t^* = \frac{1}{\sigma_V z}$ stays constant. This parameter t^* controls the temporal frequency bandwidth, as illustrated on the middle of Figure 2. We used a two alternative forced choice (2AFC) paradigm. In each trial, a gray fixation screen with a small dark fixation spot was followed by two stimulus intervals of 250 ms each, separated by an uniformly gray 250 ms inter-stimulus interval. The first stimulus had parameters (v_1, z_1) and the second had parameters (v_2, z_2) . At the end of the trial, a gray screen appeared asking the participant to report which one of the two intervals was perceived as moving faster by pressing one of two buttons, that is whether $v_1 > v_2$ or $v_2 > v_1$.

Given reference values (v^*, z^*) , for each trial, (v_1, z_1) and (v_2, z_2) are selected such that

$$\begin{cases} v_i = v^*, z_i \in z^* + \Delta_Z \\ v_j \in v^* + \Delta_V, z_j = z^* \end{cases} \quad \text{where} \quad \Delta_V = \{-2, -1, 0, 1, 2\},$$

where $(i, j) = (1, 2)$ or $(i, j) = (2, 1)$ (i.e. the ordering is randomized across trials), and where z values are expressed in cycles per degree (c°) and v values in $^\circ/s$. The range Δ_Z is defined below. Ten repetitions of each of the 25 possible combinations of these parameters are made per block of 250 trials and at least four such blocks were collected per condition tested. The outcome of these experiments are summarized by psychometric curves $\hat{\varphi}_{v^*, z^*}$, where for all $(v - v^*, z - z^*) \in \Delta_V \times \Delta_Z$, the value $\hat{\varphi}_{v^*, z^*}(v, z)$ is the empirical probability (each averaged over the typically 40 trials) that a stimulus generated with parameters (v^*, z) is moving faster than a stimulus with parameters (v, z^*) .

To assess the validity of our model, we tested different scenarios summarized in Table 1. Each row corresponds to 35 minutes of testing per participant and was always performed by at least two of the participants. Stimuli were generated on a Mac running OS 10.6.8 and displayed on a 20" Viewsonic p227f monitor with resolution 1024×768 at 100 Hz. Routines were written using Matlab 7.10.0 and Psychtoolbox 3.0.9 controlled the stimulus display. Observers sat 57 cm from the screen in a dark room. Four observers, three male and one female, with normal or corrected to normal vision

Case	t^*	σ_Z	B_Z	v^*	z^*	Δ_Z
A1	200 ms	1.0 c/°	×	5 °/s	0.8 c/°	$\{-0.27, -0.16, 0, 0.27, 0.48\}$
A2	200 ms	1.0 c/°	×	5 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
A3	200 ms	1.0 c/°	×	10 °/s	0.8 c/°	$\{-0.27, -0.16, 0, 0.27, 0.48\}$
A4	200 ms	1.0 c/°	×	10 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
B1	100 ms	1.0 c/°	×	10 °/s	0.8 c/°	$\{-0.27, -0.16, 0, 0.27, 0.48\}$
B2	100 ms	1.0 c/°	×	10 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
C1	100 ms	×	1.28	5 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
C2	100 ms	×	1.28	10 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
C3	200 ms	×	1.28	5 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$
C4	200 ms	×	1.28	10 °/s	1.28 c/°	$\{-0.48, -0.21, 0, 0.32, 0.85\}$

Table 1: A and B are both bandwidth controlled in °/s with high and low t^* respectively, C is bandwidth controlled in octaves.

took part in these experiments. They gave their informed consent and the experiments received ethical approval from the Aix-Marseille Ethics Committee in accordance with the declaration of Helsinki.

To increase the statistical power of the data set during analysis, psychometric functions were generated following the observed effect in the data and a sampling was carried out to obtain a synthetic data set for the validation of the Bayesian fitting procedure. The steps involved are detailed in section 4.5.

4.2 Bayesian modeling

To make full use of our MC paradigm in analyzing the obtained results, we follow the methodology of the Bayesian observer used for instance in [45, 44, 25]. We assume the observer makes its decision using a Maximum A Posteriori (MAP) estimator

$$\hat{v}_z(m) = \underset{v}{\operatorname{argmin}} [-\log(\mathbb{P}_{M|V,Z}(m|v, z)) - \log(\mathbb{P}_{V|Z}(v|z))] \quad (28)$$

computed from some internal representation $m \in \mathbb{R}$ of the observed stimulus. For simplicity, we assume that the observer estimates z from m without bias. To simplify the numerical analysis, we assume that the likelihood is Gaussian, with a variance independent of v . Furthermore, we assume that the prior is Laplacian as this gives a good description of the a priori statistics of speeds in natural images [10]:

$$\mathbb{P}_{M|V,Z}(m|v, z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{|m-v|^2}{2\sigma_z^2}} \quad \text{and} \quad \mathbb{P}_{V|Z}(v|z) \propto e^{a_z v} 1_{[0, v_{\max}]}(v). \quad (29)$$

where $v_{\max} > 0$ is a cutoff speed ensuring that $\mathbb{P}_{V|Z}$ is a well defined density even if $a_z > 0$.

Both a_z and σ_z are unknown parameters of the model, and are obtained from the outcome of the experiments by a fitting process we now explain.

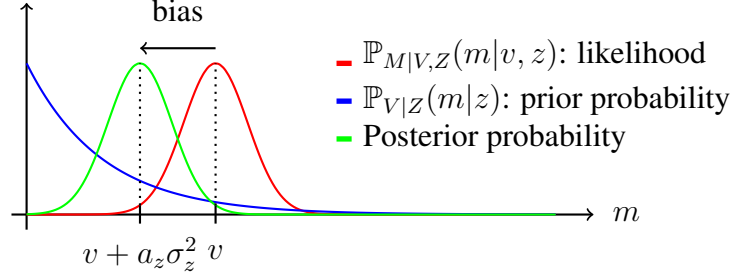


Figure 5: Multiplying the likelihood by such a prior gives a posterior that looks like a shifted version of the likelihood. Such an idea shows that the prior is responsible for bias when a Bayesian inference is performed.

4.3 Experimental Likelihood vs. the MC Model

We propose to directly fit the likelihood $\mathbb{P}_{M|V,Z}(m|v, z)$ from the experimental psychophysical curve. While this makes sense from a data-analysis point of view, this required strong modeling hypothesis, in particular, that the likelihood is Gaussian with a variance σ_z^2 independent of the parameter v to be estimated by the observer.

In this section, we derive a likelihood model directly from the stimuli, by assuming another hypothesis, that the observer uses a standard motion estimation process, based on the motion energy concept [2], an idea we incorporate here into the MC distribution. In this setting, this corresponds to using a MLE estimator, and making use of the sPDE formulation of MC.

MLE Speed Estimator. We first show how to compute this MLE estimator. To be able to achieve this, we use the sPDE formulation provided by Proposition 5. Equation (19) is useful from a Bayesian modeling perspective, because, informally, it can be interpreted as the fact that the Gaussian distribution of MC has the following appealing form, for any video $\mathcal{I} : \Omega \times T \rightarrow \mathbb{R}$ observed on a bounded space-time domain $\Omega \times [0, T]$,

$$-\log(\mathbb{P}_I(\mathcal{I}|v_0)) = Z_I + \int_{\Omega} \int_0^T |\mathcal{D}(K_W \star \mathcal{I})(x, t) + \langle \mathcal{G}(K_W \star \mathcal{I})(x, t), v_0 \rangle + \langle \mathcal{H}(K_W \star \mathcal{I})(x, t)v_0, v_0 \rangle|^2 dt dx \quad (30)$$

where K_W is the spatial filter corresponding to the square-root inverse of the covariance Σ_W , i.e. which satisfies $\hat{K}_W(\xi) \stackrel{\text{def}}{=} \hat{\sigma}_W(\xi)^{-1}$, where \mathcal{D} is defined in (16), \mathcal{G} and \mathcal{H} are defined in (20), where Z_I is a normalization constant which is independent of v_0 where $\hat{\sigma}_W$ is defined in (18). Equation (30) can be seen as a direct generalization of the initial energy model (3), when the first order luminance conservation sPDE (2) is replaced by the second order MC sPDE model (19).

It is however important to realize that the expression (30) is only formal, since the rigorous definition of the likelihood of infinite dimensional Gaussian distribution is more involved [20]. It is possible to give a simple rigorous expression for the case of discretized clouds satisfying the AR(2) recursion (26). In this case, for some input video $\mathcal{I} = (\mathcal{I}^{(\ell)})_{\ell=1}^L$, the log-likelihood reads

$$-\log(\mathbb{P}_I(\mathcal{I})) = \tilde{Z}_I + K_{v_0}(\mathcal{I}) \quad \text{where}$$

$$K_{v_0}(\mathcal{I}) \stackrel{\text{def.}}{=} \frac{1}{\Delta^4} \sum_{\ell=1}^L \int_{\Omega} |K_W \star \mathcal{I}^{(\ell+1)}(x) - \mathcal{U}_{v_0} \star K_W \star \mathcal{I}^{(\ell)}(x) - \mathcal{V}_{v_0} \star K_W \star \mathcal{I}^{(\ell-1)}(x)|^2 dx$$

where \mathcal{U}_{v_0} and \mathcal{V}_{v_0} are defined in (27). This convenient formulation can be used to re-write the MLE estimator of the horizontal speed v parameter of a MC as

$$\hat{v}^{\text{MLE}}(\mathcal{I}) \stackrel{\text{def.}}{=} \underset{v}{\operatorname{argmax}} \mathbb{P}_I(\mathcal{I}) = \underset{v}{\operatorname{argmin}} K_{v_0}(\mathcal{I}) \quad \text{where} \quad v_0 = (v, 0) \in \mathbb{R}^2 \quad (31)$$

where we used the fact that \tilde{Z}_I is independent of v_0 . The solution to this optimization problem with respect to v is then computed using the Newton-CG optimization method implemented in the python library `scipy`.

MLE Modeling of the Likelihood. Following several previous works such as [45, 44], we assumed the existence of an internal representation parameter m , which was assumed to be a scalar, with a Gaussian distribution conditioned on (v, z) . We explore here the possibility that this internal representation could be directly obtained from the stimuli by the observer using an “optimal” speed detector (an MLE estimate).

Denoting $I_{v,z}$ a MC, which is a random Gaussian field of power spectrum (8), with central speeds $v_0 = (v, 0)$ and central spatial frequency z (the other parameters being fixed as explained in the experimental section of the paper), this means that we consider the internal representation as being the following scalar random variable

$$M_{v,z} \stackrel{\text{def.}}{=} \hat{v}_z^{\text{MLE}}(I_{v,z}) \quad \text{where} \quad \hat{v}_z^{\text{MLE}}(\mathcal{I}) \stackrel{\text{def.}}{=} \underset{v}{\operatorname{argmax}} \mathbb{P}_{M|V,Z}(\mathcal{I}|v, z), \quad (32)$$

which corresponds to the optimization (31) and can be solved efficiently numerically.

As shown in Figure 6(a), we observed that $M_{v,z}$ is well approximated by a Gaussian random variable. Its mean is very close to v , and Figure 6(b) shows the evolution of its variance for different spatial frequencies z . An important point to note here is that this optimal estimation model (using an MLE) is not consistent with the experimental finding because the estimated standard deviations of observers do not show a decreasing behavior as in Figure 6(b).

4.4 Likelihood and Prior Estimation

Adopting an approach from previous literature [45, 44, 25], the theoretical psychophysical curve obtained by a Bayesian decision model is

$$\varphi_{v^*, z^*}(v, z) \stackrel{\text{def.}}{=} \mathbb{E}(\hat{v}_{z^*}(M_{v, z^*}) > \hat{v}_z(M_{v^*, z}))$$

where $M_{v,z} \sim \mathcal{N}(v, \sigma_z^2)$ is a Gaussian variable having the distribution $\mathbb{P}_{M|V,Z}(\cdot|v, z)$.

The following proposition shows that in our special case of Gaussian prior and Laplacian likelihood, it can be computed in closed form. Its proof follows closely the derivation of [44, Appendix A].

Proposition 8. *In the special case of the estimator (28) with a parameterization (29), one has*

$$\varphi_{v^*, z^*}(v, z) = \psi \left(\frac{v - v^* - a_{z^*} \sigma_{z^*}^2 + a_z \sigma_z^2}{\sqrt{\sigma_{z^*}^2 + \sigma_z^2}} \right) \quad (33)$$

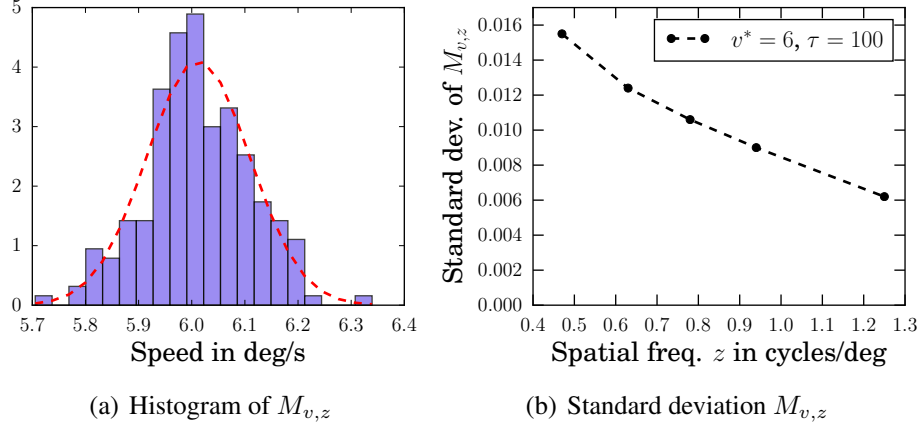


Figure 6: Estimates of $M_{v,z}$ for $z = 0.8$ c/° defined by (32) and its standard deviation as a function of z .

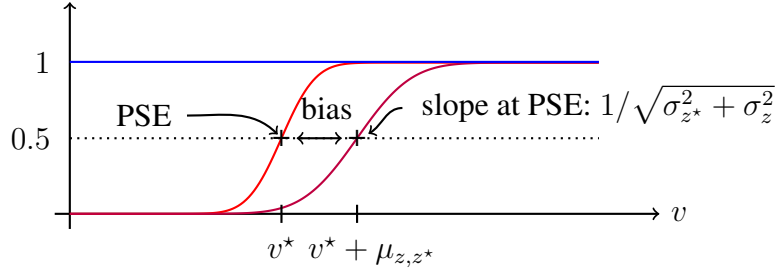


Figure 7: The shape of the psychometric function follows the estimation of the two speeds by Bayesian inference 5. This figure illustrates Proposition 8. The bias ensues from the difference between the bias on the two estimated speeds.

where $\psi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-s^2/2} ds$ is the cumulative normal function of sigmoid shape.

Proof. One has the closed form expression for the MAP estimator

$$\hat{v}_z(m) = m - a_z \sigma_z^2,$$

and hence, denoting $\mathcal{N}(\mu, \sigma^2)$ the Gaussian distribution of mean μ and variance σ^2 ,

$$\hat{v}_z(M_{v,z}) \sim \mathcal{N}(v - a_z \sigma_z^2, \sigma_z^2)$$

where \sim means equality of distributions. One thus has

$$\hat{v}_{z*}(M_{v,z*}) - \hat{v}_z(M_{v*,z}) \sim \mathcal{N}(v - v^* - a_{z*} \sigma_{z*}^2 + a_z \sigma_z^2, \sigma_{z*}^2 + \sigma_z^2),$$

which leads to the results by taking expectation. \square

Fitting procedure In order to fit this model to our data we use a two-step method each consisting in minimizing the Kullback-Leibler divergence between the model and

its samples. Numerically, the Nelder-Mead simplex method implemented in the python library `scipy` has been used. Before going further let us introduce

$$\varphi_{v^*, z^*}^{a, \sigma}(v, z) = \psi \left(\frac{v - v^* - a_{z^*} \sigma_{z^*}^2 + a_z \sigma_z^2}{\sqrt{\sigma_{z^*}^2 + \sigma_z^2}} \right),$$

$$\varphi_{v^*, z^*}^{\mu, \Sigma}(v, z) = \psi \left(\frac{v - v^* + \mu_{z^*, z}}{\Sigma_{z^*, z}} \right)$$

$$\text{and } \text{KL}(\hat{p}|p) = \hat{p} \log \left(\frac{\hat{p}}{p} \right) + (1 - \hat{p}) \log \left(\frac{1 - \hat{p}}{1 - p} \right)$$

where $\mu_{z^*, z} = a_z \sigma_z^2 - a_{z^*} \sigma_{z^*}^2$, $\Sigma_{z^*, z}^2 = \sigma_{z^*}^2 + \sigma_z^2$ and KL is the Kullback-Leibler divergence between samples \hat{p} and model p .

- Step 1: for all z, z^* , initialize at a random point, compute

$$(\hat{\mu}, \hat{\Sigma}) = \underset{\mu, \Sigma}{\text{argmin}} \sum_v \text{KL}(\hat{\varphi}_{v^*, z^*} | \varphi_{v^*, z^*}^{\mu, \Sigma})$$

- Step 2: solve the linear relation shown above between $(\hat{\mu}, \hat{\Sigma})$ and $(\hat{a}, \hat{\sigma})$
- Step 3: initialize at $(\hat{a}, \hat{\sigma})$, compute

$$(\hat{a}, \hat{\sigma}) = \underset{a, \sigma}{\text{argmin}} \sum_{z, z^*} \sum_v \text{KL}(\hat{\varphi}_{v^*, z^*} | \varphi_{v^*, z^*}^{a, \sigma})$$

Remark 3. This method is coupled with a repeated stochastic initialization for the first step in order to overcome the number of local minima encountered during the fitting process. The approach was found to exhibit better results than a direct and global fit (third point). The potential problem of KL fits producing misleading results after convergence to local minima made it necessary to extend the empirical data by generating synthetic analogous data from the psychometric fits. Through this process detailed in Section 4.5 a more robust test of the validity of the analysis can be carried out.

Remark 4. Note that in practice we perform a fit in a log-speed domain *ie* we consider $\varphi_{\tilde{v}^*, z^*}(\tilde{v}, z)$ where $\tilde{v} = \ln(1 + v/v_0)$ with $v_0 = 0.3^\circ/\text{s}$ following [45].

4.5 Synthetic Extension to Empirical Data

To avoid the dangerous aspect of undefined local minima convergence during KL fitting to empirical data, the quality of fitting can be assessed more objectively on derived synthetic data. The parameters a_z and σ_z were chosen so that they reproduce the increasing behavior of $\mu_{z^*, z} = a_z \sigma_z^2 - a_{z^*} \sigma_{z^*}^2$. Then, the values of the psychometric functions $\varphi_{v^*, z^*}^{a, \sigma}(v, z)$ at the experimental points (v_1, z_1) and (v_2, z_2) described in Section 4.1 and rows (A1) and (A2) of Table 1 were used as the parameters of a binomial distribution from which we can generate any number n_b of blocks of 10 repetitions. The ten corresponding psychometric curves are shown in Figure 8 along with their

fitted version. Following the fitting procedure described above in 4.4, we show in Figures 9 and 10 our results for $(\hat{a}, \hat{\sigma})$ and $(\hat{\hat{a}}, \hat{\hat{\sigma}})$. The quality of fitting naturally increases with the number of blocks, this effect is most striking for the likelihood width. The fitted log-prior slope shows a significant offset that is due to the under determination of the linear relations between $(\hat{\mu}, \hat{\Sigma})$ and $(\hat{a}, \hat{\sigma})$. Indeed solutions of the associated linear system lies in one dimensional affine space. However, even though the true values of a_z remain intractable the decreasing behavior of a is well captured within the trends generated by the synthetic data sets and by implication the same trends are valid in the empirical data.

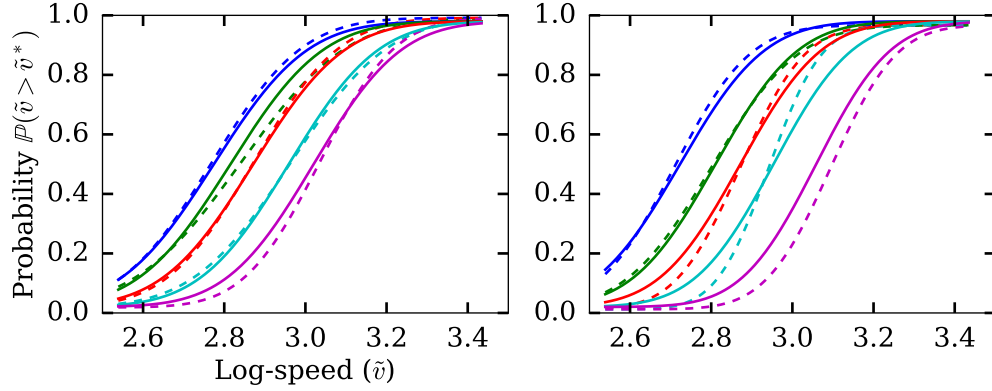


Figure 8: On the left the psychometric curves that simulate case A1, on the right the psychometric curves that simulate A2. Simulated psychometric curves resulting from the synthetic data are represented by the plain lines and the empirically fitted psychophysical curves are represented by the dotted lines.

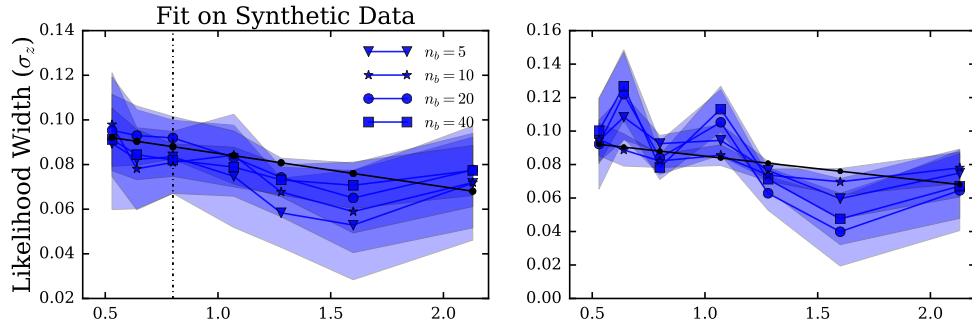


Figure 9: On the left the likelihood width $\hat{\sigma}$ obtained after the first optimization step 4.4, on the right the likelihood width $\hat{\hat{\sigma}}$ obtained after the third optimization step 4.4. These estimations are represented for different numbers of block with one standard deviation error. The black line represents the ground truth values of the likelihood widths used to generate the synthetic data.

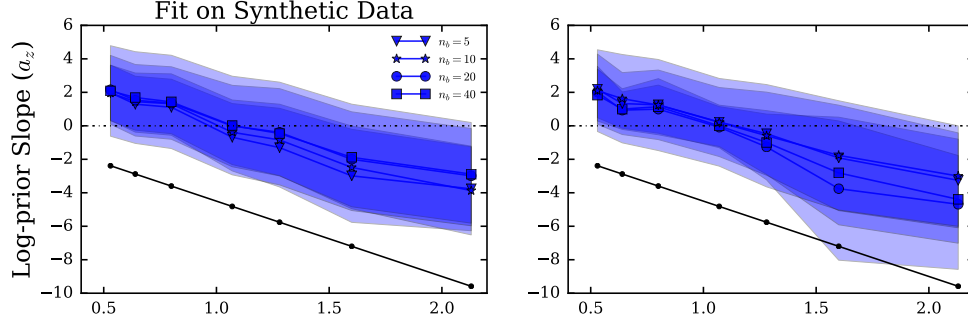


Figure 10: On the left the prior slope \hat{a} obtained after the first optimization step 4.4, on the right the prior slope $\hat{\hat{a}}$ obtained after the third optimization step 4.4. These estimations are represented for different numbers of block with one standard deviation error. The black line represents the ground truth values of the prior slopes used to generate the synthetic data.

4.6 Results

Estimating speed in dynamic visual scenes is undoubtedly a crucial skill for the successful interaction of any animal with its environment. Human judgements of perceived speed have therefore generated much interest, and been studied with a range psychophysics paradigms. The different results obtained in these studies suggest that rather than computing a veridical estimate, the visual system generates speed judgements influenced by contrast [46], speed range [47], luminance [22], spatial frequency [6, 40, 42] and retinal eccentricity [23]. There are currently no theoretical models of the underlying mechanisms serving speed estimation which capture this dependence on such a broad range of image characteristics. One of the reasons might be that the simplified grating stimuli used in most of the previous studies do not shed light on the possible elaborations in neural processing that arise when more complex stimulation. Such elaborations, such as nonlinearities in spatio-temporal frequency space can be seen in their simplest form even with a superposition of a pair gratings [36]. In the current work, we used our formulation of motion cloud stimuli which allowed the separate parametric manipulation of peak spatial frequency (z), spatial frequency bandwidth (B_z, σ_z) and stimulus lifetime (t^*) which is inversely related to the temporal variability. The stimuli are all broadband, closer resembling visual inputs under natural stimulation. In the plotted data, we avoid cluttering by restricting traces to a subset of data, S1/S2, from the pair of participants who completed the full set of parametric conditions. Our approach was to test fewer participants (4) but under several parametric conditions using a large number trials analyzed alongside the synthetic data. The data that is not plotted here shows trends that lie within the range of patterns seen from S1/S2.

Cycle-controlled bandwidth conditions The main manipulation in each case was the direct comparison of the speed of a range of five stimuli in which the central spatial frequency was varied between five values, but all other parameters were equated under the different conditions. In a first manipulation in which bandwidth was controlled by fixing it at a value of 1 c/° for all stimuli (conditions A* and B* in Table 1), we found that lower frequencies were consistently perceived to be moving slower than higher fre-

quencies (see Figure 11). The bias was generally smaller at 5 °/s than at 10 °/s (compare first column on the left with remaining two columns). This trend was the same for both the lower and the higher spatial frequency ranges used in the tasks (see Table 1 for details) when we compare the top row, Figure 11(a) with the bottom row, Figure 11(b). This means the effect generalizes across the two scales used. The temporal variability of the stimulus manipulated via t^* was found to increase the variability of the bias estimates, though this did not significantly increase the biases (compare the shaded errors in the pair of plots in both the second and the third columns of Figure 11).

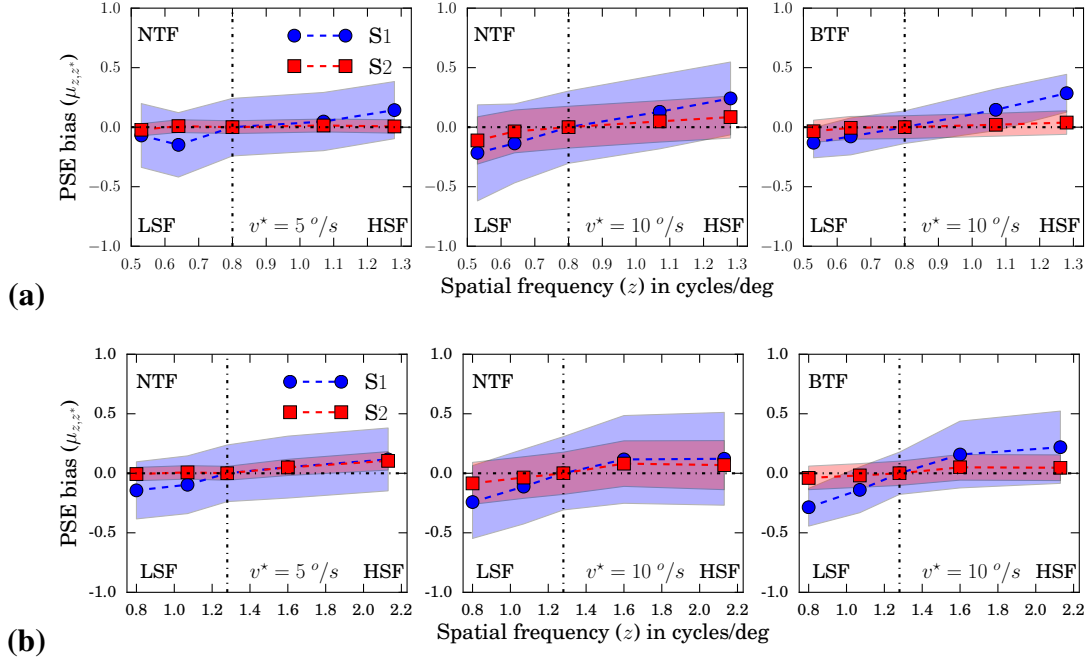


Figure 11: *Relative perceived speeds from the Point of Subjective Equality (PSE).* (a) From left to right A1, A3, B1. (b) From left to right A2, A4, B2. Task generates psychometric functions which show shifts in the point of subjective equality for the range of test z . Stimuli of lower frequency with respect to the reference (intersection of dotted horizontal and vertical lines gives the reference stimulus) are perceived as going slower, those with greater mean frequency are perceived as going relatively faster. This effect is observed under all conditions but is stronger for subject 1. Error bars are computed from those obtained for $(\hat{a}_z, \hat{\sigma}_z)$ which explains their amplitude. In case of a direct fitting of μ_{z,z^*} they are significantly smaller (not shown).

Octave-controlled bandwidth conditions The octave-bandwidth controlled stimuli of conditions C* (see Table 1), allowed us to vary the spatial frequency manipulations (z) in a way that generated scale invariant bandwidths exactly as would be expected from zooming movements towards or away from scene objects (see Figure 1). Thus if trends seen in Figure 11 were the result of ecologically invalid fixing of bandwidths at 1 c/° in the manipulations, this would be corrected in the current manipulation. Only the

higher frequency comparison range was used. We found that the trend was the same as that seen in Figure 11, indeed higher spatial frequencies were consistently perceived as faster than lower ones, shown in Figure 12. Interestingly, for the bandwidth controlled stimuli, the biases do not change across speed conditions (compare left column with right hand side columns of Figure 12). A small systematic change in the bias is seen with the manipulation of t^* , reducing temporal variability going from the upper to the lower row reduces the measured biases. The bias at the highest frequency averaged for S1/S2 is equal to 0.13 for $t^* = 100$ ms (BTF) and equal to 0.08 for $t^* = 200$ ms (NTF).

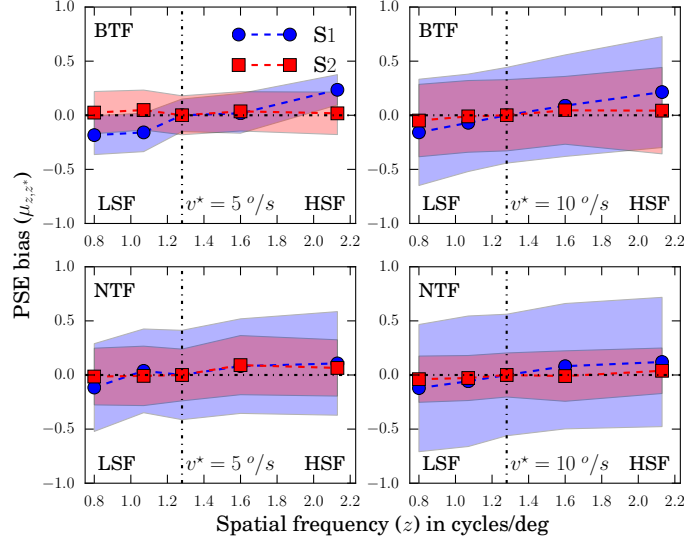


Figure 12: *Relative perceived speeds from the Point of Subjective Equality (PSE).* Top: C1, C2. Bottom: C3, C4. Same comment as Figure 11. The effect does not appear for subject 2 in case C2 and C3. Error bars are computed from those obtained for $(\hat{a}_z, \hat{\sigma}_z)$ which explains their amplitude. In case of a direct fitting of μ_{z, z^*} they are significantly smaller (not shown).

Measured biases and corresponding sensory likelihoods and priors We used the Bayesian formulation detailed in Section 4.4 to estimate the likelihood widths and the corresponding prior slopes under the tested experimental conditions. There is no systematic trend within the likelihoods in the cycle-bandwidth controlled condition fits in Figure 13(a) and there is also individual variability in the trends. We conclude that the sensory variability of the speed estimates obtained from the Bayesian modeling cannot explain the spatial frequency driven bias in perceived speed that is measured. The log prior slopes show a systematic reduction as spatial frequency is increased, see in Figure 13(b). Under all conditions, the data is best explained by a decreasing log prior as spatial frequencies are increasing. Under the octave-bandwidth controlled stimulus condition, the trends in changes in the best fitted likelihoods as the spatial frequency is increased are again not systematic (Figure 14(a)). The log prior slopes do however show a small systematic reduction as spatial frequencies are increased, in Figure 14(b).

The slopes are less steep than under the cycle-bandwidth manipulations (linear regression gives an average of -2.08 for the log-prior slopes in Figure 13(b) and -1.31 for the log-prior slopes in Figure 14(b)). Under both bandwidth configurations, we conclude that the prior slope explains at least part of the systematic effect of spatial frequency on perceived speed.

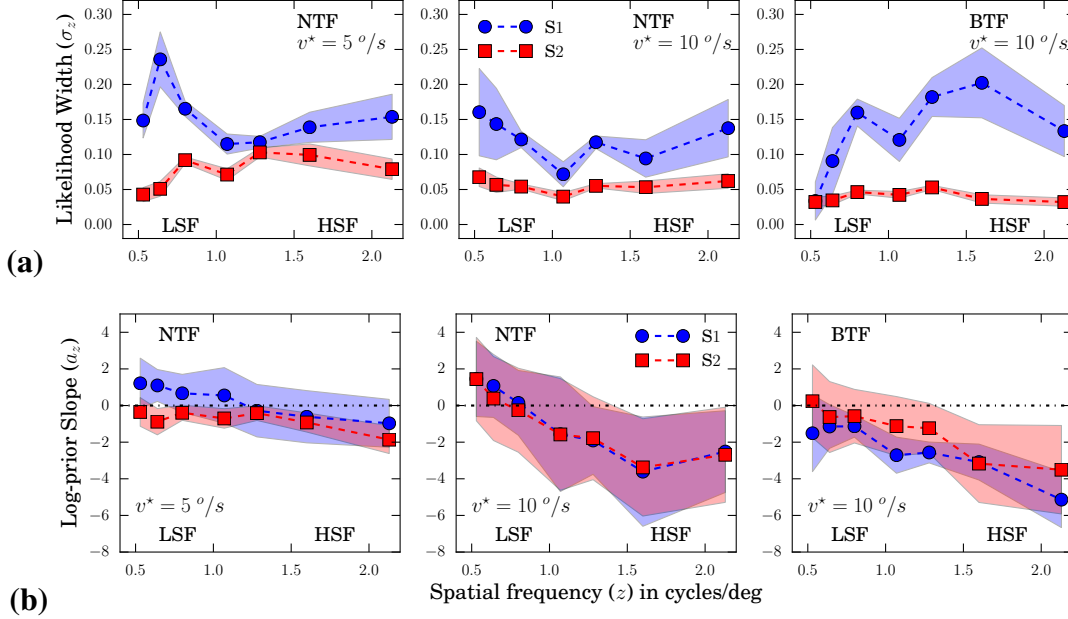


Figure 13: Likelihood widths and log-prior slopes. (a) Likelihood widths for A1-A2, A3-A4 and B1-B2. Likelihood widths do not show any common behavior, different behavior are observed for subject 1 whereas it is almost constant for subject 2. (b) Log prior slopes for A1-A2, A3-A4 and B1-B2. Despite the amplitude of error bars the log prior slopes have a common decreasing behavior in all subjects and in all cases.

4.7 Insights into Human Speed Perception

We exploited the principled and ecologically motivated parameterization of MC to study biases in human speed judgements under a range of parametric conditions. Primarily, we considered the effect of scene scaling on perceived speed, manipulated via central spatial frequencies in a similar way to previous experiments which had shown spatial frequency induced perceived speed biases [6, 41]. In general, our experimental result confirmed that higher spatial frequencies were consistently perceived to be moving faster than compared lower frequencies; the same result reported in a previous study using both simple gratings and compounds of paired gratings, the second of which can be considered as a relatively broadband bandwidth stimulus [6]. In that work, they noted that biases were present, but slightly reduced in the compound (broadband) stimuli. That conclusion was consistent with a more recent psychophysics manipulation in which up to four distinct composite gratings were used in relative speed judge-

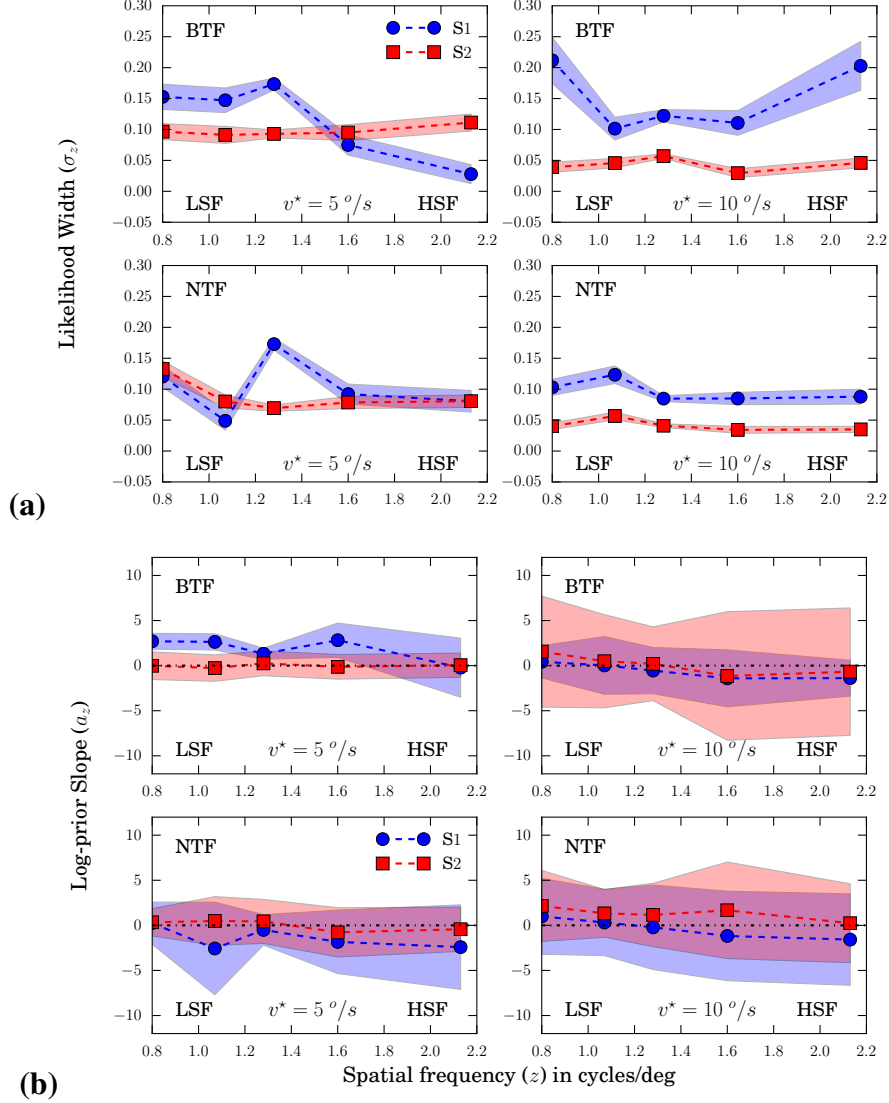


Figure 14: Likelihood widths and log-prior slopes. (a) Likelihood widths. Top: C1, C2. Bottom: C3, C4. Same as Figure 13(a). (b) Log prior slopes. Top: C1, C2. Bottom: C3, C4. Same as Figure 13(b) except for subject 2 in case C3.

ments. Estimates were found to be more veridical as bandwidth increased by adding additional components from the set of four, but increasing spatial frequencies generally biased towards faster perceived speed even if individual participants showed different trends [25]. Indeed, findings from primate neurophysiology studies have also noted that while responses are biased by spatial frequency, the tendency towards true speed sensitivity (measured as the proportion of individual neurons showing speed sensitivity) increases when broadband stimulation is used [36, 35].

It is increasingly being recognized that linear systems approaches to interrogating visual processing with single sinusoidal luminance grating inputs represents a powerful, but limited, approach to studying speed perception as they fail to capture the fact that naturalistic broadband frequency distributions may support speed estimation [6, 31, 32]. A linear consideration for example would not account for the fact that estimation

in the presence of multiple sinusoidal components results in non-linear optimal combination [25]. The current work sought to extend the body of previous work by looking at spatial frequency induced biases using a parametric configuration in the form of the motion clouds which allowed a manipulation across a continuous scale of frequency and bandwidth parameters. The effect of frequency interactions across the broadband stimulus defined along the two dimensional orthogonal spatio-temporal luminance plane to allowed us to measure the perceptual effect of the projection of different areas (e.g. see Figure 2) onto the same speed line. The measurement should rely on proposed inhibitory interactions which occur during spatio-temporal frequency integration for speed perception [40] which cannot be seen with component stimuli separated by octaves [25].

We used a faster and slower speed because previous work using sinusoidal grating stimuli had shown that below the slower range ($< 8^\circ/\text{s}$), uncertainty manipulated through lower contrasts caused an under estimation of speeds while at faster speeds ($> 16^\circ/\text{s}$) it caused an overestimation [47, 22]. Our findings show that under the cycle-controlled bandwidth conditions, biases were larger at the faster speed than the slower ones while under the octave controlled bandwidths, the biases were almost identical for both speeds. The projections made from the frequency plane onto the speed line at these two speeds, once corrected with a scale invariance assumption, was therefore the same at these two speeds which typically show differences in contrast manipulations. Indeed the Bayesian fitting did not identify a systematic shift of either likelihood or prior slope parameters that could explain the biases observed particularly for the bandwidth controlled condition. While the current work does not resolve the ongoing gaps in our understanding of speed perception mechanisms particularly as it did not tackle contrast related biases, it showed that known frequency biases in speed perception also arise from orthogonal spatial and temporal uncertainties when RMS contrast is controlled. Bayesian models such as the one we applied, which effectively project distributions in the spatiotemporal plane onto a given speed line in which a linear low speed prior applies [45] may be insufficient to capture the actual spatiotemporal priors. Indeed the Bayesian models which successfully predict speed perception with more complex or composite stimuli often require various elaborations away from simplistic low speed priors [25, 44]. Indeed even imaging studies considering the underlying mechanisms fail to find definitive evidence for the encoding of a slow speed prior [53].

5 Conclusions

We have proposed and detailed a generative model for the estimation of the motion of dynamic images based on a formalization of small perturbations from the observer’s point of view during parameterized rotations, zooms and translations. We connected these transformations to descriptions of ecologically motivated movements of both observers and the dynamic world. The fast synthesis of naturalistic textures optimized to probe motion perception was then demonstrated, through fast GPU implementations applying auto-regression techniques with much potential for future experimentation. This extends previous work from [39] by providing an axiomatic formulation. Finally, we used the stimuli in a psychophysical task and showed that these textures allow one

to further understand the processes underlying speed estimation. We used broadband stimulation to study frequency induced biases in visual perception, using various stimulus configuration including octave bandwidth and RMS contrast controlled manipulations which allowed us to manipulate central frequencies as scale invariant stimulus zooms. We showed that measured biases under these controlled conditions were the same at both a faster and a slower tested speed. By linking the stimulation directly to the standard Bayesian formalism, we demonstrated that the sensory representation of the stimulus (the likelihoods) in such models can be described directly from the generative MC model. The widely accepted Bayesian model which assumes a slow speed prior showed that the frequency interactions could not be fully captured by the current formulation. We conclude that an extension to that formulation is needed and perhaps a two dimensional prior acting on the frequency space and mediated by underlying neural sensitivity has a role to play in computational modeling of complex spatiotemporal integration behind speed perception. We propose that more experiments with naturalistic stimuli such as MCs and a consideration of more generally applicable priors will be needed in future.

Acknowledgments

We thank Guillaume Masson for useful discussions during the development of the experiments. We also thank Manon Bouyé and Élise Amfreville for proofreading. LUP was supported by EC FP7-269921, “BrainScaleS” and BalaV1 ANR-13-BSV4-0014-02. The work of JV and GP was supported by the European Research Council (ERC project SIGMA-Vision). AIM and LUP were supported by SPEED ANR-13-SHS2-0006.

References

- [1] B. Abraham, O. I. Camps, and M. Sznaier. “Dynamic texture with Fourier descriptors”. In: *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*. Vol. 1. 2005, pp. 53–58.
- [2] E. H. Adelson and J. R. Bergen. “Spatiotemporal energy models for the perception of motion”. In: *Journal of Optical Society of America, A*. 2.2 (Feb. 1985), pp. 284–99.
- [3] R. T. Born and D. C. Bradley. “Structure and function of visual area MT”. In: *Annual review of neuroscience* 28.1 (2005), pp. 157–89.
- [4] P. Brockwell, R. Davis, and Y. Yang. “Continuous-time Gaussian autoregression”. In: *Statistica Sinica* 17.1 (2007), p. 63.
- [5] P. J. Brockwell and A. Lindner. “Existence and uniqueness of stationary Lévy-driven CARMA processes”. In: *Stochastic Processes and their Applications* 119.8 (2009), pp. 2660–2681.

- [6] K. R. Brooks, T. Morris, and P. Thompson. “Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: Implications for MT models of speed perception”. In: *Journal of vision* 11.14 (2011), pp. 19–19.
- [7] M. Colombo and P. Seriès. “Bayes in the brain: on Bayesian modelling in neuroscience”. In: *The British journal for the philosophy of science* 63.3 (2012), pp. 697–723.
- [8] R. Costantini, L. Sbaiz, and S. Süsstrunk. “Higher order SVD analysis for dynamic texture synthesis”. In: *Image Processing, IEEE Transactions on* 17.1 (2008), pp. 42–52.
- [9] J. J. J. DiCarlo, D. Zoccolan, and N. C. C. Rust. “How Does the Brain Solve Visual Object Recognition?” In: *Neuron* 73.3 (2012), pp. 415–434.
- [10] D. Dong. “Maximizing Causal Information of Natural Scenes in Motion”. In: *Dynamics of Visual Motion Processing*. Springer US, 2010, pp. 261–282.
- [11] G. Doretto et al. “Dynamic Textures”. In: *International Journal of Computer Vision* 51.2 (Feb. 2003), pp. 91–109.
- [12] K. Doya. *Bayesian brain: Probabilistic approaches to neural coding*. MIT press, 2007.
- [13] N. El Karoui, S. Peng, and M. C. Quenez. “Backward stochastic differential equations in finance”. In: *Mathematical finance* 7.1 (1997), pp. 1–71.
- [14] J. Filip, M. Haindl, and D. Chetverikov. “Fast synthesis of dynamic colour textures”. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. Vol. 4. IEEE. 2006, pp. 25–28.
- [15] R. F. Fox. “Stochastic versions of the Hodgkin-Huxley equations”. In: *Biophysical Journal* 72.5 (1997), pp. 2068–2074.
- [16] B. Galerne. “Stochastic image models and texture synthesis”. PhD thesis. ENS de Cachan, 2011.
- [17] B. Galerne, Y. Gousseau, and J. M. Morel. “Micro-Texture Synthesis by Phase Randomization”. In: *Image Processing On Line* 1 (2011).
- [18] B. Galerne, Y. Gousseau, and J. M. Morel. “Random Phase Textures: Theory and Synthesis.” In: *IEEE T. Image. Process.* (2010).
- [19] I. M. Gel’fand, N. Y. Vilenkin, and A. Feinstein. “Generalized functions. Vol. 4.” In: (1964).
- [20] E. Giné and R. Nickl. “Mathematical foundations of infinite-dimensional statistical models”. In: *Cambridge Series in Statistical and Probabilistic Mathematics* (2015).
- [21] R. L. Gregory. “Perceptions as hypotheses”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 290.1038 (July 1980), pp. 181–197.
- [22] O. Hassan and S. T. Hammett. “Perceptual biases are inconsistent with Bayesian encoding of speed in the human visual system”. In: *Journal of vision* 15.2 (2015), pp. 9–9.

- [23] O. Hassan, P. Thompson, and S. T. Hammett. “Perceived speed in peripheral vision can go up or down”. In: *Journal of vision* 16.6 (2016), pp. 20–20.
- [24] M. Hyndman, A. D. Jepson, and D. J. Fleet. “Higher-order Autoregressive Models for Dynamic Textures.” In: *BMVC*. 2007, pp. 1–10.
- [25] M. Jogan and A. A. Stocker. “Signal Integration in Human Visual Speed Perception”. In: *The Journal of Neuroscience* 35.25 (2015), pp. 9381–9390.
- [26] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions, vol. 1-2*. New York: John Wiley & Sons, 1994, pp. 211–213.
- [27] D. Kersten, P. Mamassian, and A. Yuille. “Object perception as Bayesian inference”. In: *Annu. Rev. Psychol.* 55 (2004), pp. 271–304.
- [28] D. C. Knill and A. Pouget. “The Bayesian brain: the role of uncertainty in neural coding and computation”. In: *TRENDS in Neurosciences* 27.12 (2004), pp. 712–719.
- [29] C.-B. Liu et al. “Dynamic Textures Synthesis as Nonlinear Manifold Learning and Traversing.” In: *BMVC*. Citeseer. 2006, pp. 859–868.
- [30] R. Meidan. “On the connection between ordinary and generalized stochastic processes”. In: *Journal of Mathematical Analysis and Applications* 76.1 (1980), pp. 124–133.
- [31] A. I. Meso and C. Simoncini. “Towards an understanding of the roles of visual areas MT and MST in computing speed”. In: *Frontiers in computational neuroscience* 8 (2014).
- [32] A. I. Meso and J. M. Zanker. “Speed encoding in correlation motion detectors as a consequence of spatial structure”. In: *Biological cybernetics* 100.5 (2009), pp. 361–370.
- [33] O. Nestares, D. Fleet, and D. Heeger. “Likelihood functions and confidence bounds for total-least-squares problems”. In: *IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000*. Vol. 1. Hilton Head Island, SC, USA: IEEE Comput. Soc, 2000, pp. 523–530.
- [34] F. Oberhettinger. *Tables of Mellin transforms*. Springer Science & Business Media, 2012.
- [35] J. A. Perrone and A. Thiele. “Speed skills: measuring the visual speed analyzing properties of primate MT neurons”. In: *Nat. Neurosci.* 4.5 (May 2001), pp. 526–532.
- [36] N. Priebe, C. Cassanello, and S. Lisberger. “The neural representation of speed in macaque area MT/V5.” In: *J. Neurosci.* 23 (2003), pp. 5650–5661.
- [37] A. Rahman, M. Murshed, et al. “Dynamic texture synthesis using motion distribution statistics”. In: *Journal of Research and Practice in Information Technology* 40.2 (2008), p. 129.
- [38] G. A. Rousselet, S. J. Thorpe, and M. Fabre-Thorpe. “How parallel is visual processing in the ventral pathway?” In: *Trends in Cognitive Sciences* 8.8 (2004), pp. 363–370.

- [39] P. Sanz-Leon et al. “Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception”. In: *Journal of Neurophysiology* 107.11 (Mar. 2012), pp. 3217–3226.
- [40] C. Simoncini et al. “More is not always better: adaptive gain control explains dissociation between perception and action”. In: *Nature Neurosci* 15.11 (Nov. 2012), pp. 1596–1603.
- [41] A. T. Smith and G. K. Edgar. “The influence of spatial frequency on perceived temporal frequency and perceived speed.” In: *Vision Res.* 30 (1990), pp. 1467–1474.
- [42] M. A. Smith, N. Majaj, and J. A. Movshon. “Dynamics of Pattern Motion Computation”. In: *Dynamics of Visual Motion Processing: Neuronal, Behavioral and Computational Approaches*. Ed. by G. S. Masson and U. J. Ilg. First. Berlin-Heidelberg: Springer, 2010, pp. 55–72.
- [43] P. L. Smith. “Stochastic dynamic models of response time and accuracy: A foundational primer”. In: *Journal of mathematical psychology* 44.3 (2000), pp. 408–463.
- [44] G. Sotiropoulos, A. R. Seitz, and P. Seriès. “Contrast dependency and prior expectations in human speed perception”. In: *Vision Research* 97.0 (2014), pp. 16–23.
- [45] A. A. Stocker and E. P. Simoncelli. “Noise characteristics and prior expectations in human visual speed perception”. In: *Nature Neuroscience* 9.4 (Mar. 2006), pp. 578–585.
- [46] P. Thompson. “Perceived rate of movement depends on contrast”. In: *Vision research* 22.3 (1982), pp. 377–380.
- [47] P. Thompson, K. Brooks, and S. T. Hammett. “Speed can go up as well as down at low contrast: Implications for models of motion perception”. In: *Vision research* 46.6 (2006), pp. 782–786.
- [48] M. Unser et al. “A Unified Formulation of Gaussian Versus Sparse Stochastic Processes – Part II: Discrete-Domain Theory”. In: *IEEE Transactions on Information Theory* 60.5 (2014), pp. 3036–3051.
- [49] M. Unser and P. Tafti. *An Introduction to Sparse Stochastic Processes*. 367 p. Cambridge, UK: Cambridge University Press, 2014.
- [50] M. Unser, P. D. Tafti, and Q. Sun. “A unified formulation of Gaussian versus sparse stochastic processes – Part I: Continuous-domain theory”. In: *Information Theory, IEEE Transactions on* 60.3 (2014), pp. 1945–1962.
- [51] J. Vacher et al. “Biologically Inspired Dynamic Textures for Probing Motion Perception”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 1909–1917.
- [52] N. G. Van K. *Stochastic processes in physics and chemistry*. Vol. 1. Elsevier, 1992.

- [53] B. Vintch and J. L. Gardner. “Cortical correlates of human motion perception biases”. In: *The Journal of Neuroscience* 34.7 (2014), pp. 2592–2604.
- [54] L. Y. Wei et al. “State of the Art in Example-based Texture Synthesis”. In: *Eurographics 2009, State of the Art Report, EG-STAR*. Eurographics Association, 2009.
- [55] X. X. Wei and A. A. Stocker. “Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference”. In: *NIPS*. Ed. by P. L. Bartlett et al. 2012, pp. 1313–1321.
- [56] Y. Weiss and D. J. Fleet. “Velocity likelihoods in biological and machine vision”. In: *In Probabilistic Models of the Brain: Perception and Neural Function*. 2001, pp. 81–100.
- [57] Y. Weiss, E. P. Simoncelli, and E. H. Adelson. “Motion illusions as optimal percepts”. In: *Nature Neuroscience* 5.6 (June 2002), pp. 598–604.
- [58] G. S. Xia et al. “Synthesizing and Mixing Stationary Gaussian Texture Models”. In: *SIAM Journal on Imaging Sciences* 7.1 (2014), pp. 476–508.
- [59] R. A. Young and R. M. L. “The gaussian derivative model for spatial-temporal vision: II. Cortical data”. In: *Spatial vision* 14.3 (2001), pp. 321–390.
- [60] L. Yuan et al. “Synthesizing dynamic texture with closed-loop linear dynamic system”. In: *Computer Vision-ECCV 2004*. Springer, 2004, pp. 603–616.